(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 January 2002 (31.01.2002)

PCT

(10) International Publication Number
WO 02/08396 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/US01/22397

(22) International Filing Date: 17 July 2001 (17.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|---|---|---|
| 60/220,063 | 21 July 2000 (21.07.2000) | US |
| 60/221,680 | 28 July 2000 (28.07.2000) | US |
| 60/223,544 | 4 August 2000 (04.08.2000) | US |
| 60/224,717 | 11 August 2000 (11.08.2000) | US |
| 60/225,988 | 16 August 2000 (16.08.2000) | US |
| 60/227,568 | 23 August 2000 (23.08.2000) | US |

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DELEGEANE, Angelo, M.** [US/US]; 594 Angus Drive, Milpitas, CA 95035 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue, 1, Menlo Park, CA 94025 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way, #1, Menlo Park, CA 94025 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee, #5, San Francisco, CA 94107 (US). **DAS, Debopriya** [IN/US]; 1179 Bonita Avenue, #3, Mountain View, CA 94040 (US). **KALLICK, Deborah, A.** [US/US]; 900 Olive Street, Mento Park, CA 94025 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **KHAN, Farrah, A.** [IN/US]; 3617 Central Road, Glenview, IL 60025 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **POLICKY, Jennifer, L.** [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **YANG, Junming** [CN/US]; 7125 Bank Lane, San Jose, CA 95129 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue, #23, Mountain View, CA 94043 (US). **DING, Li** [CN/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US). **KEARNEY, Liam** [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, Sunnyvale, CA 94577 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **SANJANWALA, Madhu, S.** [US/US]; 210 Sylvia Court, Los Altos, CA 94024 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street, #205, San Leandro, CA 94577 (US). **LAL, Preeti** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **LEE, Sally** [US/US]; 825 East Evelyn, #425, Sunnyvale, CA 94086 (US). **TODD, Stephen** [US/US]; 2239 45th Avenue, San Francisco, CA 94116 (US). **LO, Terrence, P.** [CA/US]; 1451 Beach Park Boulevard, #115, Foster City, CA 94404 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polten Place Way, San Jose, CA 95121 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *without international search report and to be republished upon receipt of that report*

*[Continued on next page]*

(54) Title: PROTEASES

(57) Abstract: The invention provides human proteases (PRTS) and polynucleotides which identify and encode PRTS. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PRTS.

WO 02/08396 A2

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# PROTEASES

## TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of proteases and to the use of these sequences in the diagnosis, treatment, and prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

## BACKGROUND OF THE INVENTION

Proteases cleave proteins and peptides at the peptide bond that forms the backbone of the protein or peptide chain. Proteolysis is one of the most important and frequent enzymatic reactions that occurs both within and outside of cells. Proteolysis is responsible for the activation and maturation of nascent polypeptides, the degradation of misfolded and damaged proteins, and the controlled turnover of peptides within the cell. Proteases participate in digestion, endocrine function, and tissue remodeling during embryonic development, wound healing, and normal growth. Proteases can play a role in regulatory processes by affecting the half life of regulatory proteins. Proteases are involved in the etiology or progression of disease states such as inflammation, angiogenesis, tumor dispersion and metastasis, cardiovascular disease, neurological disease, and bacterial, parasitic, and viral infections.

Proteases can be categorized on the basis of where they cleave their substrates. Exopeptidases, which include aminopeptidases, dipeptidyl peptidases, tripeptidases, carboxypeptidases, peptidyl-di-peptidases, dipeptidases, and omega peptidases, cleave residues at the termini of their substrates. Endopeptidases, including serine proteases, cysteine proteases, and metalloproteases, cleave at residues within the peptide. Four principal categories of mammalian proteases have been identified based on active site structure, mechanism of action, and overall three-dimensional structure. (See Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 1-5.)

Serine Proteases

The serine proteases (SPs) are a large, widespread family of proteolytic enzymes that include the digestive enzymes trypsin and chymotrypsin, components of the complement and blood-clotting cascades, and enzymes that control the degradation and turnover of macromolecules within the cell and in the extracellular matrix. Most of the more than 20 subfamilies can be grouped into six clans,

each with a common ancestor. These six clans are hypothesized to have descended from at least four evolutionarily distinct ancestors. SPs are named for the presence of a serine residue found in the active catalytic site of most families. The active site is defined by the catalytic triad, a set of conserved asparagine, histidine, and serine residues critical for catalysis. These residues form a charge relay network that facilitates substrate binding. Other residues outside the active site form an oxyanion hole that stabilizes the tetrahedral transition intermediate formed during catalysis. SPs have a wide range of substrates and can be subdivided into subfamilies on the basis of their substrate specificity. The main subfamilies are named for the residue(s) after which they cleave: trypases (after arginine or lysine), aspases (after aspartate), chymases (after phenylalanine or leucine), metases (methionine), and serases (after serine) (Rawlings, N.D. and A.J. Barrett (1994) Methods Enzymol. 244:19-61).

Most mammalian serine proteases are synthesized as zymogens, inactive precursors that are activated by proteolysis. For example, trypsinogen is converted to its active form, trypsin, by enteropeptidase. Enteropeptidase is an intestinal protease that removes an N-terminal fragment from trypsinogen. The remaining active fragment is trypsin, which in turn activates the precursors of the other pancreatic enzymes. Likewise, proteolysis of prothrombin, the precursor of thrombin, generates three separate polypeptide fragments. The N-terminal fragment is released while the other two fragments, which comprise active thrombin, remain associated through disulfide bonds.

The two largest SP subfamilies are the chymotrypsin (S1) and subtilisin (S8) families. Some members of the chymotrypsin family contain two structural domains unique to this family. Kringle domains are triple-looped, disulfide cross-linked domains found in varying copy number. Kringles are thought to play a role in binding mediators such as membranes, other proteins or phospholipids, and in the regulation of proteolytic activity (PROSITE PDOC00020). Apple domains are 90 amino-acid repeated domains, each containing six conserved cysteines. Three disulfide bonds link the first and sixth, second and fifth, and third and fourth cysteines (PROSITE PDOC00376). Apple domains are involved in protein-protein interactions. S1 family members include trypsin, chymotrypsin, coagulation factors IX-XII, complement factors B, C, and D, granzymes, kallikrein, and tissue- and urokinase-plasminogen activators. The subtilisin family has members found in the eubacteria, archaebacteria, eukaryotes, and viruses. Subtilisins include the proprotein-processing endopeptidases kexin and furin and the pituitary prohormone convertases PC1, PC2, PC3, PC6, and PACE4 (Rawlings and Barrett, supra).

SPs have functions in many normal processes and some have been implicated in the etiology or treatment of disease. Enterokinase, the initiator of intestinal digestion, is found in the intestinal

brush border, where it cleaves the acidic propeptide from trypsinogen to yield active trypsin (Kitamoto, Y. et al. (1994) Proc. Natl. Acad. Sci. USA 91:7588-7592). Prolylcarboxypeptidase, a lysosomal serine peptidase that cleaves peptides such as angiotensin II and III and [des-Arg9] bradykinin, shares sequence homology with members of both the serine carboxypeptidase and prolylendopeptidase families (Tan, F. et al. (1993) J. Biol. Chem. 268:16631-16638). The protease neuropsin may influence synapse formation and neuronal connectivity in the hippocampus in response to neural signaling (Chen, Z.-L. et al. (1995) J. Neurosci. 15:5088-5097). Tissue plasminogen activator is useful for acute management of stroke (Zivin, J.A. (1999) Neurology 53:14-19) and myocardial infarction (Ross, A.M. (1999) Clin. Cardiol. 22:165-171). Some receptors (PAR, for proteinase-activated receptor), highly expressed throughout the digestive tract, are activated by proteolytic cleavage of an extracellular domain. The major agonists for PARs, thrombin, trypsin, and mast cell tryptase, are released in allergy and inflammatory conditions. Control of PAR activation by proteases has been suggested as a promising therapeutic target (Vergnolle, N. (2000) Aliment. Pharmacol. Ther. 14:257-266; Rice, K.D. et al. (1998) Curr. Pharm. Des. 4:381-396). Prostate-specific antigen (PSA) is a kallikrein-like serine protease synthesized and secreted exclusively by epithelial cells in the prostate gland. Serum PSA is elevated in prostate cancer and is the most sensitive physiological marker for monitoring cancer progression and response to therapy. PSA can also identify the prostate as the origin of a metastatic tumor (Brawer, M.K. and P.H. Lange (1989) Urology 33:11-16).

The signal peptidase is a specialized class of SP found in all prokaryotic and eukaryotic cell types that serves in the processing of signal peptides from certain proteins. Signal peptides are amino-terminal domains of a protein which direct the protein from its ribosomal assembly site to a particular cellular or extracellular location. Once the protein has been exported, removal of the signal sequence by a signal peptidase and posttranslational processing, e.g., glycosylation or phosphorylation, activate the protein. Signal peptidases exist as multi-subunit complexes in both yeast and mammals. The canine signal peptidase complex is composed of five subunits, all associated with the microsomal membrane and containing hydrophobic regions that span the membrane one or more times (Shelness, G.S. and G. Blobel (1990) J. Biol. Chem. 265:9512-9519). Some of these subunits serve to fix the complex in its proper position on the membrane while others contain the actual catalytic activity.

Another family of proteases which have a serine in their active site are dependent on the hydrolysis of ATP for their activity. These proteases contain proteolytic core domains and regulatory ATPase domains which can be identified by the presence of the P-loop, an ATP/GTP-binding motif (PROSITE PDOC00803). Members of this family include the eukaryotic mitochondrial matrix proteases, Clp protease and the proteasome. Clp protease was originally found in plant chloroplasts

but is believed to be widespread in both prokaryotic and eukaryotic cells. The gene for early-onset torsion dystonia encodes a protein related to Clp protease (Ozelius, L.J. et al. (1998) Adv. Neurol. 78:93-105).

The proteasome is an intracellular protease complex found in some bacteria and in all eukaryotic cells, and plays an important role in cellular physiology. Proteasomes are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins of all types, including proteins that function to activate or repress cellular processes such as transcription and cell cycle progression (Ciechanover, A. (1994) Cell 79:13-21). In the UCS pathway, proteins targeted for degradation are conjugated to ubiquitin, a small heat stable protein. The ubiquitinated protein is then recognized and degraded by the proteasome. The resultant ubiquitin-peptide complex is hydrolyzed by a ubiquitin carboxyl terminal hydrolase, and free ubiquitin is released for reutilization by the UCS. Ubiquitin-proteasome systems are implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes (p53), cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, supra). This pathway has been implicated in a number of diseases, including cystic fibrosis, Angelman's syndrome, and Liddle syndrome (reviewed in Schwartz, A.L. and A. Ciechanover (1999) Annu. Rev. Med. 50:57-74). A murine proto-oncogene, Unp, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH3T3 cells. The human homologue of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) Oncogene 10:2179-2183). Ubiquitin carboxyl terminal hydrolase is involved in the differentiation of a lymphoblastic leukemia cell line to a non-dividing mature state (Maki, A. et al. (1996) Differentiation 60:59-66). In neurons, ubiquitin carboxyl terminal hydrolase (PGP 9.5) expression is strong in the abnormal structures that occur in human neurodegenerative diseases (Lowe, J. et al. (1990) J. Pathol. 161:153-160). The proteasome is a large (~2000 kDa) multisubunit complex composed of a central catalytic core containing a variety of proteases arranged in four seven-membered rings with the active sites facing inwards into the central cavity, and terminal ATPase subunits covering the outer port of the cavity and regulating substrate entry (for review, see Schmidt, M. et al. (1999) Curr. Opin. Chem. Biol. 3:584-591).

Cysteine Proteases

Cysteine proteases (CPs) are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Nearly half of the CPs known are present only in viruses. CPs have a cysteine as the major catalytic residue at the active site where catalysis proceeds via a thioester intermediate and is facilitated by nearby histidine and asparagine

residues. A glutamine residue is also important, as it helps to form an oxyanion hole. Two important CP families include the papain-like enzymes (C1) and the calpains (C2). Papain-like family members are generally lysosomal or secreted and therefore are synthesized with signal peptides as well as propeptides. Most members bear a conserved motif in the propeptide that may have structural significance (Karrer, K.M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:3063-3067). Three-dimensional structures of papain family members show a bilobed molecule with the catalytic site located between the two lobes. Papains include cathepsins B, C, H, L, and S, certain plant allergens and dipeptidyl peptidase (for a review, see Rawlings, N.D. and A.J. Barrett (1994) Methods Enzymol. 244:461-486).

Some CPs are expressed ubiquitously, while others are produced only by cells of the immune system. Of particular note, CPs are produced by monocytes, macrophages and other cells which migrate to sites of inflammation and secrete molecules involved in tissue repair. Overabundance of these repair molecules plays a role in certain disorders. In autoimmune diseases such as rheumatoid arthritis, secretion of the cysteine peptidase cathepsin C degrades collagen, laminin, elastin and other structural proteins found in the extracellular matrix of bones. Bone weakened by such degradation is also more susceptible to tumor invasion and metastasis. Cathepsin L expression may also contribute to the influx of mononuclear cells which exacerbates the destruction of the rheumatoid synovium (Keyszer, G.M. (1995) Arthritis Rheum. 38:976-984).

Calpains are calcium-dependent cytosolic endopeptidases which contain both an N-terminal catalytic domain and a C-terminal calcium-binding domain. Calpain is expressed as a proenzyme heterodimer consisting of a catalytic subunit unique to each isoform and a regulatory subunit common to different isoforms. Each subunit bears a calcium-binding EF-hand domain. The regulatory subunit also contains a hydrophobic glycine-rich domain that allows the enzyme to associate with cell membranes. Calpains are activated by increased intracellular calcium concentration, which induces a change in conformation and limited autolysis. The resultant active molecule requires a lower calcium concentration for its activity (Chan, S.L. and M.P. Mattson (1999) J. Neurosci. Res. 58:167-190). Calpain expression is predominantly neuronal, although it is present in other tissues. Several chronic neurodegenerative disorders, including ALS, Parkinson's disease and Alzheimer's disease are associated with increased calpain expression (Chan and Mattson, supra). Calpain-mediated breakdown of the cytoskeleton has been proposed to contribute to brain damage resulting from head injury (McCracken, E. et al. (1999) J. Neurotrauma 16:749-761). Calpain-3 is predominantly expressed in skeletal muscle, and is responsible for limb-girdle muscular dystrophy type 2A (Minami, N. et al. (1999) J. Neurol. Sci. 171:31-37).

Another family of thiol proteases is the caspases, which are involved in the initiation and execution phases of apoptosis. A pro-apoptotic signal can activate initiator caspases that trigger a proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death of the cell. Two active site residues, a cysteine and a histidine, have been implicated in the catalytic mechanism. Caspases are among the most specific endopeptidases, cleaving after aspartate residues. Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain interacts with cofactors that can positively or negatively affect apoptosis. An activating signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1 numbering convention) and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long prodomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. Some caspases contain a "death effector domain" in their prodomain by which they can be recruited into self-activating complexes with other caspases and FADD protein associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer. Endogenous caspase inhibitors (inhibitor of apoptosis proteins, or IAPs) also exist. All these interactions have clear effects on the control of apoptosis (reviewed in Chan and Mattson, supra; Salveson, G.S. and V.M. Dixit (1999) Proc. Natl. Acad. Sci. USA 96:10964-10967).

Caspases have been implicated in a number of diseases. Mice lacking some caspases have severe nervous system defects due to failed apoptosis in the neuroepithelium and suffer early lethality. Others show severe defects in the inflammatory response, as caspases are responsible for processing IL-1b and possibly other inflammatory cytokines (Chan and Mattson, supra). Cowpox virus and baculoviruses target caspases to avoid the death of their host cell and promote successful infection. In addition, increases in inappropriate apoptosis have been reported in AIDS, neurodegenerative diseases and ischemic injury, while a decrease in cell death is associated with cancer (Salveson and Dixit, supra; Thompson, C.B. (1995) Science 267:1456-1462).

Aspartyl proteases

Aspartyl proteases (APs) include the lysosomal proteases cathepsins D and E, as well as chymosin, renin, and the gastric pepsins. Most retroviruses encode an AP, usually as part of the pol polyprotein. APs, also called acid proteases, are monomeric enzymes consisting of two domains, each domain containing one half of the active site with its own catalytic aspartic acid residue. APs are most active in the range of pH 2–3, at which one of the aspartate residues is ionized and the other

neutral. The pepsin family of APs contains many secreted enzymes, and all are likely to be synthesized with signal peptides and propeptides. Most family members have three disulfide loops, the first ~5 residue loop following the first aspartate, the second 5-6 residue loop preceding the second aspartate, and the third and largest loop occurring toward the C terminus. Retropepsins, on the other hand, are analogous to a single domain of pepsin, and become active as homodimers with each retropepsin monomer contributing one half of the active site. Retropepsins are required for processing the viral polyproteins.

APs have roles in various tissues, and some have been associated with disease. Renin mediates the first step in processing the hormone angiotensin, which is responsible for regulating electrolyte balance and blood pressure (reviewed in Crews, D.E. and S.R. Williams (1999) Hum. Biol. 71:475-503). Abnormal regulation and expression of cathepsins are evident in various inflammatory disease states. Expression of cathepsin D is elevated in synovial tissues from patients with rheumatoid arthritis and osteoarthritis. The increased expression and differential regulation of the cathepsins are linked to the metastatic potential of a variety of cancers (Chambers, A.F. et al. (1993) Crit. Rev. Oncol. 4:95-114).

Metalloproteases

Metalloproteases require a metal ion for activity, usually manganese or zinc. Examples of manganese metalloenzymes include aminopeptidase P and human proline dipeptidase (PEPD). Aminopeptidase P can degrade bradykinin, a nonapeptide activated in a variety of inflammatory responses. Aminopeptidase P has been implicated in coronary ischemia/reperfusion injury. Administration of aminopeptidase P inhibitors has been shown to have a cardioprotective effect in rats (Ersahin, C. et al (1999) J. Cardiovasc. Pharmacol. 34:604-611).

Most zinc-dependent metalloproteases share a common sequence in the zinc-binding domain. The active site is made up of two histidines which act as zinc ligands and a catalytic glutamic acid C-terminal to the first histidine. Proteins containing this signature sequence are known as the metzincins and include aminopeptidase N, angiotensin-converting enzyme, neurolysin, the matrix metalloproteases and the adamalysins (ADAMS). An alternate sequence is found in the zinc carboxypeptidases, in which all three conserved residues – two histidines and a glutamic acid – are involved in zinc binding.

A number of the neutral metalloendopeptidases, including angiotensin converting enzyme and the aminopeptidases, are involved in the metabolism of peptide hormones. High aminopeptidase B activity, for example, is found in the adrenal glands and neurohypophyses of hypertensive rats (Prieto, I. et al. (1998) Horm. Metab. Res. 30:246-248). Oligopeptidase M/neurolysin can hydrolyze bradykinin as well as neurotensin (Serizawa, A. et al. (1995) J. Biol. Chem 270:2092-2098).

Neurotensin is a vasoactive peptide that can act as a neurotransmitter in the brain, where it has been implicated in limiting food intake (Tritos, N.A. et al. (1999) Neuropeptides 33:339-349).

The matrix metalloproteases (MMPs) are a family of at least 23 enzymes that can degrade components of the extracellular matrix (ECM). They are $Zn^{+2}$ endopeptidases with an N-terminal catalytic domain. Nearly all members of the family have a hinge peptide and C-terminal domain which can bind to substrate molecules in the ECM or to inhibitors produced by the tissue (TIMPs, for tissue inhibitor of metalloprotease; Campbell, I.L. et al. (1999) Trends Neurosci. 22:285). The presence of fibronectin-like repeats, transmembrane domains, or C-terminal hemopexinase-like domains can be used to separate MMPs into collagenase, gelatinase, stromelysin and membrane-type MMP subfamilies. In the inactive form, the $Zn^{+2}$ ion in the active site interacts with a cysteine in the pro-sequence. Activating factors disrupt the $Zn^{+2}$-cysteine interaction, or "cysteine switch," exposing the active site. This partially activates the enzyme, which then cleaves off its propeptide and becomes fully active. MMPs are often activated by the serine proteases plasmin and furin. MMPs are often regulated by stoichiometric, noncovalent interactions with inhibitors; the balance of protease to inhibitor, then, is very important in tissue homeostasis (reviewed in Yong, V.W. et al. (1998) Trends Neurosci. 21:75).

MMPs are implicated in a number of diseases including osteoarthritis (Mitchell, P. et al. (1996) J. Clin. Invest. 97:761), atherosclerotic plaque rupture (Sukhova, G.K. et al. (1999) Circulation 99:2503), aortic aneurysm (Schneiderman, J. et al. (1998) Am. J. Path. 152:703), non-healing wounds (Saarialho-Kere, U.K. et al. (1994) J. Clin. Invest. 94:79), bone resorption (Blavier, L. and J.M. Delaisse (1995) J. Cell Sci. 108:3649), age-related macular degeneration (Steen, B. et al. (1998) Invest. Ophthalmol. Vis. Sci. 39:2194), emphysema (Finlay, G.A. et al. (1997) Thorax 52:502), myocardial infarction (Rohde, L.E. et al. (1999) Circulation 99:3063) and dilated cardiomyopathy (Thomas, C.V. et al. (1998) Circulation 97:1708). MMP inhibitors prevent metastasis of mammary carcinoma and experimental tumors in rat, and Lewis lung carcinoma, hemangioma, and human ovarian carcinoma xenografts in mice (Eccles, S.A. et al. (1996) Cancer Res. 56:2815; Anderson et al. (1996) Cancer Res. 56:715-718; Volpert, O.V. et al. (1996) J. Clin. Invest. 98:671; Taraboletti, G. et al. (1995) J. NCI 87:293; Davies, B. et al. (1993) Cancer Res. 53:2087). MMPs may be active in Alzheimer's disease. A number of MMPs are implicated in multiple sclerosis, and administration of MMP inhibitors can relieve some of its symptoms (reviewed in Yong, supra).

Another family of metalloproteases is the ADAMs, for A Disintegrin and Metalloprotease Domain, which they share with their close relatives the adamalysins, snake venom metalloproteases (SVMPs). ADAMs combine features of both cell surface adhesion molecules and proteases,

containing a prodomain, a protease domain, a disintegrin domain, a cysteine rich domain, an epidermal growth factor repeat, a transmembrane domain, and a cytoplasmic tail. The first three domains listed above are also found in the SVMPs. The ADAMs possess four potential functions: proteolysis, adhesion, signaling and fusion. The ADAMs share the metzincin zinc binding sequence and are inhibited by some MMP antagonists such as TIMP-1.

ADAMs are implicated in such processes as sperm-egg binding and fusion, myoblast fusion, and protein-ectodomain processing or shedding of cytokines, cytokine receptors, adhesion proteins and other extracellular protein domains (Schlöndorff, J. and C.P. Blobel (1999) J. Cell. Sci. 112:3603-3617). The Kuzbanian protein cleaves a substrate in the NOTCH pathway (possibly NOTCH itself), activating the program for lateral inhibition in Drosophila neural development. Two ADAMs, TACE (ADAM 17) and ADAM 10, are proposed to have analogous roles in the processing of amyloid precursor protein in the brain (Schlöndorff and Blobel, supra). TACE has also been identified as the TNF activating enzyme (Black, R.A. et al. (1997) Nature 385:729). TNF is a pleiotropic cytokine that is important in mobilizing host defenses in response to infection or trauma, but can cause severe damage in excess and is often overproduced in autoimmune disease. TACE cleaves membrane-bound pro-TNF to release a soluble form. Other ADAMs may be involved in a similar type of processing of other membrane-bound molecules. MADDAM (for metalloprotease and disintegrin dendritic antigen marker), a member of the ADAM19 family, is up-regulated in monocytes induced to become dendritic cells. It is useful as a marker for distinguishing between dendritic cells and macrophages (Fritsche, J. et al. (2000) Blood 96:732-739).

The ADAMTS sub-family has all of the features of ADAM family metalloproteases and contain an additional thrombospondin domain (TS). The prototypic ADAMTS was identified in mouse, found to be expressed in heart and kidney and upregulated by proinflammatory stimuli (Kuno, K. et al. (1997) J. Biol. Chem. 272:556-562). To date eleven members are recognized by the Human Genome Organization (HUGO; http://www.gene.ucl.ac.uk/users/hester/adamts.html#Approved). Members of this family have the ability to degrade aggrecan, a high molecular weight proteoglycan which provides cartilage with important mechanical properties including compressibility, and which is lost during the development of arthritis. Enzymes which degrade aggrecan are thus considered attractive targets to prevent and slow the degradation of articular cartilage (See, e.g., Tortorella, M.D. (1999) Science 284:1664; Abbaszade, I. (1999) J. Biol. Chem. 274:23443). Other members are reported to have antiangiogenic potential (Kuno et al., supra) and/or procollagen processing (Colige, A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2374).

The discovery of new proteases, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

## SUMMARY OF THE INVENTION

The invention features purified polypeptides, proteases, referred to collectively as "PRTS" and individually as "PRTS-1," "PRTS-2," "PRTS-3," "PRTS-4," "PRTS-5," "PRTS-6," "PRTS-7," "PRTS-8," "PRTS-9," "PRTS-10," "PRTS-11," "PRTS-12," "PRTS-13," "PRTS-14," "PRTS-15," "PRTS-16," "PRTS-17," "PRTS-18," "PRTS-19," "PRTS-20," and "PRTS-21." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-21. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:22-42.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a

biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ

ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino

acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one

test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:22-42, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a

polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

## BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

**DEFINITIONS**

"PRTS" refers to the amino acid sequences of substantially purified PRTS obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PRTS. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

An "allelic variant" is an alternative form of the gene encoding PRTS. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PRTS include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PRTS or a polypeptide with at least one functional characteristic of PRTS. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PRTS, and improper or unexpected hybridization to allelic variants, with a

locus other than the normal chromosomal locus for the polynucleotide sequence encoding PRTS. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PRTS. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PRTS is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PRTS. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PRTS polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PRTS, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PRTS or fragments of PRTS may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

| Original Residue | Conservative Substitution |
|---|---|
| Ala | Gly, Ser |
| Arg | His, Lys |
| Asn | Asp, Gln, His |
| Asp | Asn, Glu |
| Cys | Ala, Ser |
| Gln | Asn, Glu, His |
| Glu | Asp, Gln, His |
| Gly | Ala |
| His | Asn, Arg, Gln, Glu |
| Ile | Leu, Val |
| Leu | Ile, Val |
| Lys | Arg, Gln, Glu |
| Met | Leu, Ile |
| Phe | His, Met, Leu, Trp, Tyr |
| Ser | Cys, Thr |
| Thr | Ser, Val |
| Trp | Phe, Tyr |
| Tyr | His, Phe, Trp |
| Val | Ile, Leu, Thr |

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of PRTS or the polynucleotide encoding PRTS which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:22-42 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:22-42, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:22-42 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:22-42 from related polynucleotide sequences. The precise length of a fragment of SEQ ID

NO:22-42 and the region of SEQ ID NO:22-42 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-21 is encoded by a fragment of SEQ ID NO:22-42. A fragment of SEQ ID NO:1-21 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-21. For example, a fragment of SEQ ID NO:1-21 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-21. The precise length of a fragment of SEQ ID NO:1-21 and the region of SEQ ID NO:1-21 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis

programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.

Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ($T_m$) for the specific sequence at a defined ionic strength and pH. The $T_m$ is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating $T_m$ and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., $C_0t$ or $R_0t$ analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid

support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PRTS which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PRTS which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PRTS. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PRTS.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an PRTS may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PRTS.

"Probe" refers to nucleic acid sequences encoding PRTS, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase

sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent,

chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PRTS, nucleic acids encoding PRTS, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based

on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The

presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human proteases (PRTS), the polynucleotides encoding PRTS, and the use of these compositions for the diagnosis, treatment, or prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column

3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are proteases. For example, SEQ ID NO:1 is 85% identical to human calpain 3; calcium activated neutral protease (GenBank ID g7684607) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a calpain family cysteine protease domain, an EF-hand domain and a calpain large subunit, domain III as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:1 is a protease. In an alternative example, SEQ ID NO:5 is 89% identical to human ubiquitin hydrolyzing enzyme I (GenBank ID g3220154) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 also contains a ubiquitin carboxyl terminal hydrolase active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:5 is a ubiquitin protease. In another alternative example, SEQ ID NO:15 has 56% local identity to mouse mast cell metalloprotease-6 (GenBank ID g200507) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.7e-60, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:15 also contains a trypsin family serine protease active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) The presence of this domain is confirmed by BLIMPS, MOTIFS, and PROFILESCAN analyses. BLIMPS analysis also reveals the presence of kringle and type I fibronectin domains, providing further corroborative evidence that SEQ ID NO:15 is a serine protease of the trypsin family. In yet another alternative example, SEQ ID NO:17 has 36% local identity to limulus coagulation

factor C precursor (GenBank ID g217397) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.1e-53, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:17 also contains a trypsin family protease active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) This same analysis reveals the presence of CUB and EGF-like domains. Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:17 is a serine protease of the trypsin family. In still another alternative example, SEQ ID NO:18 is 93% identical to human disintegrin and metalloprotease domain 19 (GenBank ID g6651071) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:18 also contains a neutral zinc metalloprotease active site and a disintegrin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:18 is a metalloprotease of the ADAM family. In an alternative example, SEQ ID NO:20 has 73% local identity to mouse ubiquitin specific protease (GenBank ID g7673618) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:20 also contains ubiquitin carboxyl-terminal hydrolase active site domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:20 is a ubiquitin specific protease. SEQ ID NO:2-4, SEQ ID NO:6-14, SEQ ID NO:16, SEQ ID NO:19 and SEQ ID NO:21 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-21 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification

technologies that identify SEQ ID NO:22-42 or that distinguish between SEQ ID NO:22-42 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 4847254F8 is the identification number of an Incyte cDNA sequence, and SPLNTUT02 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71666762V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g7377067) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_*XXXXXX*_$N_1$_$N_2$_*YYYYY*_$N_3$_$N_4$ represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is the identification number of a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a

RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

| Prefix | Type of analysis and/or examples of programs |
|---|---|
| GNN, GFG, ENST | Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK). |
| GBI | Hand-edited analysis of genomic sequences. |
| FL | Stitched or stretched genomic sequences (see Example V). |
| INCY | Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript. |

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PRTS variants. A preferred PRTS variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PRTS amino acid sequence, and which contains at least one functional or structural characteristic of PRTS.

The invention also encompasses polynucleotides which encode PRTS. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42, which encodes PRTS. The polynucleotide sequences of SEQ ID NO:22-42, as presented in the Sequence Listing, embrace the equivalent RNA sequences,

wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PRTS. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PRTS. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:22-42. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PRTS.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PRTS, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PRTS, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PRTS and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PRTS under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PRTS or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PRTS and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PRTS and PRTS derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PRTS or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:22-42 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PRTS may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060).

Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PRTS may be cloned in recombinant DNA molecules that direct expression of PRTS, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PRTS.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PRTS-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of PRTS, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PRTS may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PRTS itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PRTS, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PRTS, the nucleotide sequences encoding PRTS or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in

a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PRTS. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PRTS. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PRTS and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PRTS and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PRTS. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola,

M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PRTS. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PRTS can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PRTS into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PRTS are needed, e.g. for the production of antibodies, vectors which direct high level expression of PRTS may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PRTS. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PRTS. Transcription of sequences encoding PRTS may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PRTS may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PRTS in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PRTS in cell lines is preferred. For example, sequences encoding PRTS can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C: and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins

(GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PRTS is inserted within a marker gene sequence, transformed cells containing sequences encoding PRTS can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PRTS under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PRTS and that express PRTS may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PRTS using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PRTS is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PRTS include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PRTS, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase

such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PRTS may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PRTS may be designed to contain signal sequences which direct secretion of PRTS through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PRTS may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PRTS protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PRTS activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site

located between the PRTS encoding sequence and the heterologous protein sequence, so that PRTS may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PRTS may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, $^{35}$S-methionine.

PRTS of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PRTS. At least one and up to a plurality of test compounds may be screened for specific binding to PRTS. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PRTS, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PRTS binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PRTS, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PRTS or cell membrane fractions which contain PRTS are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PRTS or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PRTS, either in solution or affixed to a solid support, and detecting the binding of PRTS to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PRTS of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PRTS. Such compounds may include agonists, antagonists, or partial or

inverse agonists. In one embodiment, an assay is performed under conditions permissive for PRTS activity, wherein PRTS is combined with at least one test compound, and the activity of PRTS in the presence of a test compound is compared with the activity of PRTS in the absence of the test compound. A change in the activity of PRTS in the presence of the test compound is indicative of a compound that modulates the activity of PRTS. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PRTS under conditions suitable for PRTS activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PRTS may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PRTS or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PRTS may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PRTS can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PRTS is injected into animal ES cells, and the injected sequence

integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PRTS, e.g., by secreting PRTS in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PRTS and proteases. In addition, the expression of PRTS is closely associated with neurological, cardiovascular, hemic, prostate, endocrine, reproductive, immune system, bone and tumorus tissues and Alzheimer's disease. Therefore, PRTS appears to play a role in gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders. In the treatment of disorders associated with increased PRTS expression or activity, it is desirable to decrease the expression or activity of PRTS. In the treatment of disorders associated with decreased PRTS expression or activity, it is desirable to increase the expression or activity of PRTS.

Therefore, in one embodiment, PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose

veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism,

hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic

myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia.

In another embodiment, a vector capable of expressing PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PRTS in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PRTS may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those listed above.

In a further embodiment, an antagonist of PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS. Examples of such disorders include, but are not limited to, those gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders described above. In one aspect, an antibody which specifically binds PRTS may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PRTS.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate

therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PRTS may be produced using methods which are generally known in the art. In particular, purified PRTS may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PRTS. Antibodies to PRTS may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PRTS or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PRTS have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PRTS amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PRTS may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate

antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PRTS-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PRTS may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PRTS and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PRTS epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PRTS. Affinity is expressed as an association constant, $K_a$, which is defined as the molar concentration of PRTS-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The $K_a$ determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PRTS epitopes, represents the average affinity, or avidity, of the antibodies for PRTS. The $K_a$ determined for a preparation of monoclonal antibodies, which are monospecific for a particular PRTS epitope, represents a true measure of affinity. High-affinity antibody preparations with $K_a$ ranging from about $10^9$ to $10^{12}$ L/mole are preferred for use in immunoassays in which the PRTS-

antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with $K_a$ ranging from about $10^6$ to $10^7$ L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PRTS, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PRTS-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PRTS, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PRTS. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PRTS. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PRTS may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PRTS expression or regulation causes disease, the expression of PRTS from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PRTS are treated by constructing mammalian expression vectors encoding PRTS and introducing these vectors by mechanical means into PRTS-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of PRTS include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PRTS may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the

tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PRTS from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J..1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PRTS expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PRTS under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et

al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PRTS to cells which have one or more genetic abnormalities with respect to the expression of PRTS. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PRTS to target cells which have one or more genetic abnormalities with respect to the expression of PRTS. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PRTS to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PRTS to target cells. The biology of the prototypic alphavirus,

Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PRTS into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PRTS-coding RNAs and the synthesis of high levels of PRTS in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PRTS into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PRTS.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides,

corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PRTS. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PRTS. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PRTS expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PRTS may be therapeutically useful, and in the treatment of disorders associated with decreased PRTS expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PRTS may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method

commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PRTS is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PRTS are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PRTS. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient.

Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PRTS, antibodies to PRTS, and mimetics, agonists, antagonists, or inhibitors of PRTS.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PRTS or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PRTS or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PRTS or fragments thereof, antibodies of PRTS, and agonists, antagonists or inhibitors of PRTS, which

ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the $ED_{50}$ (the dose therapeutically effective in 50% of the population) or $LD_{50}$ (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the $LD_{50}/ED_{50}$ ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the $ED_{50}$ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 $\mu$g to 100,000 $\mu$g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind PRTS may be used for the diagnosis of disorders characterized by expression of PRTS, or in assays to monitor patients being treated with PRTS or agonists, antagonists, or inhibitors of PRTS. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PRTS include methods which utilize the antibody and a label to detect PRTS in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PRTS, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PRTS expression. Normal or standard values for PRTS expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PRTS under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PRTS expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PRTS may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PRTS may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PRTS, and to monitor regulation of PRTS levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PRTS or closely related molecules may be used to identify nucleic acid sequences which encode PRTS. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PRTS, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PRTS encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:22-42 or from genomic sequences including promoters, enhancers, and introns of the PRTS gene.

Means for producing specific hybridization probes for DNAs encoding PRTS include the cloning of polynucleotide sequences encoding PRTS or PRTS derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as $^{32}P$ or $^{35}S$, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PRTS may be used for the diagnosis of disorders associated with expression of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis,

myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of

Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding PRTS may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PRTS expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PRTS may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PRTS may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PRTS in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PRTS, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PRTS, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PRTS may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide

encoding PRTS, or a fragment of a polynucleotide complementary to the polynucleotide encoding PRTS, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PRTS include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify

genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PRTS, fragments of PRTS, or antibodies specific for PRTS may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties.

These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or

untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PRTS to quantify the levels of PRTS expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized

by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PRTS may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PRTS on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse,

may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PRTS, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PRTS and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PRTS, or fragments thereof, and washed. Bound PRTS is then detected by methods well known in the art. Purified PRTS can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PRTS specifically compete with a test compound for binding PRTS. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PRTS.

In additional embodiments, the nucleotide sequences which encode PRTS may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/220,063, U.S. Ser. No. 60/221,680, U.S. Ser. No. 60/223,544, U.S. Ser. No. 60/224,717, U.S. Ser. No. 60/225,988, and U.S. Ser. No. 60/227,568 are expressly incorporated by reference herein.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells

including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

**II. Isolation of cDNA Clones**

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

**III. Sequencing and Analysis**

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and

programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID

NO:22-42. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

**IV. Identification and Editing of Coding Sequences from Genomic DNA**

Putative proteases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode proteases, the encoded polypeptides were analyzed by querying against PFAM models for proteases. Potential proteases were also identified by homology to Incyte cDNA sequences that had been annotated as proteases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

**V. Assembly of Genomic Sequence Data with cDNA Sequence Data**

**"Stitched" Sequences**

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on

more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

**VI. Chromosomal Mapping of PRTS Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:22-42 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:22-42 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences

had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:37 was mapped to chromosome 17 within the interval from 69.3 to 74.5 centiMorgans, and to chromosome 23 within the interval from 68.2 to 90.8 centiMorgans. Similarly, SEQ ID NO:32 was mapped to chromosome 16 within the interval from 81.8 to 84.4 centiMorgans. Additionally, SEQ ID NO:31 was mapped to chromosome 3 within the interval from 88.2 to 90.1 centiMorgans, and within the interval from 91.0 to 97.2 centiMorgans. More than one map location is reported for SEQ ID NO:37 and SEQ ID NO:31, indicating that sequences having different map locations were assembled into a single cluster. This situation occurs, for example, when sequences having strong similarity, but not complete identity, are assembled into a single cluster.

**VII. Analysis of Polynucleotide Expression**

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{\text{5 x minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PRTS are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PRTS. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

## VIII. Extension of PRTS Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing $Mg^{2+}$, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 $\mu$l to 10 $\mu$l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended

clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

**IX. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:22-42 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 $\mu$Ci of [$\gamma$-$^{32}$P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing $10^7$ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate.

Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

## X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

**Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription

from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

**Microarray Preparation**

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

**Hybridization**

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with

an 1.8 $cm^2$ coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

**Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and

measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

**XI. Complementary Polynucleotides**

Sequences complementary to the PRTS-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PRTS. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PRTS. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PRTS-encoding transcript.

**XII. Expression of PRTS**

Expression and purification of PRTS is achieved using bacterial or virus-based expression systems. For expression of PRTS in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PRTS upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PRTS in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PRTS by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PRTS is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PRTS at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PRTS obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, XVIII, and XIX where applicable.

**XIII. Functional Assays**

PRTS function is assessed by expressing the sequences encoding PRTS at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 $\mu$g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 $\mu$g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of

fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PRTS on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PRTS and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PRTS and other genes of interest can be analyzed by northern analysis or microarray techniques.

**XIV. Production of PRTS Specific Antibodies**

PRTS substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PRTS amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PRTS activity by, for example, binding the peptide or PRTS to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

**XV. Purification of Naturally Occurring PRTS Using Specific Antibodies**

Naturally occurring or recombinant PRTS is substantially purified by immunoaffinity chromatography using antibodies specific for PRTS. An immunoaffinity column is constructed by covalently coupling anti-PRTS antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PRTS are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PRTS (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PRTS binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PRTS is collected.

**XVI. Identification of Molecules Which Interact with PRTS**

PRTS, or biologically active fragments thereof, are labeled with $^{125}$I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PRTS, washed, and any wells with labeled PRTS complex are assayed. Data obtained using different concentrations of PRTS are used to calculate values for the number, affinity, and association of PRTS with the candidate molecules.

Alternatively, molecules interacting with PRTS are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PRTS may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

**XVII. Demonstration of PRTS Activity**

Protease activity is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York, NY, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases, or metalloproteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (carboxypeptidases A and B, procollagen C-proteinase). Commonly used chromogens are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. Assays are performed at ambient temperature and contain an aliquot of the enzyme and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette, and the increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate is measured. The change in absorbance is proportional to the enzyme activity in the assay.

In the alternative, an assay for protease activity takes advantage of fluorescence resonance energy transfer (FRET) that occurs when one donor and one acceptor fluorophore with an appropriate spectral overlap are in close proximity. A flexible peptide linker containing a cleavage site specific for PRTS is fused between a red-shifted variant (RSGFP4) and a blue variant (BFP5) of Green Fluorescent Protein. This fusion protein has spectral properties that suggest energy transfer is occurring from BFP5 to RSGFP4. When the fusion protein is incubated with PRTS, the substrate is cleaved, and the two fluorescent proteins dissociate. This is accompanied by a marked decrease in energy transfer which is quantified by comparing the emission spectra before and after the addition of PRTS (Mitra, R.D. et al (1996) Gene 173:13-17). This assay can also be performed in living cells. In this case the fluorescent substrate protein is expressed constitutively in cells and PRTS is introduced on an inducible vector so that FRET can be monitored in the presence and absence of PRTS (Sagot, I. et al (1999) FEBS Letters 447:53-57).

**XVIII. Identification of PRTS Substrates**

Phage display libraries can be used to identify optimal substrate sequences for PRTS. A random hexamer followed by a linker and a known antibody epitope is cloned as an N-terminal extension of gene III in a filamentous phage library. Gene III codes for a coat protein, and the epitope will be displayed on the surface of each phage particle. The library is incubated with PRTS under proteolytic conditions so that the epitope will be removed if the hexamer codes for a PRTS cleavage site. An antibody that recognizes the epitope is added along with immobilized protein A. Uncleaved phage, which still bear the epitope, are removed by centrifugation. Phage in the supernatant are then amplified and undergo several more rounds of screening. Individual phage clones are then isolated and sequenced. Reaction kinetics for these peptide substrates can be studied using an assay in Example XVII, and an optimal cleavage sequence can be derived (Ke, S.H. et al. (1997) J. Biol. Chem. 272:16603-16609).

To screen for in vivo PRTS substrates, this method can be expanded to screen a cDNA expression library displayed on the surface of phage particles (T7SELECT™10-3 Phage display vector, Novagen, Madison, WI) or yeast cells (pYD1 yeast display vector kit; Invitrogen, Carlsbad, CA). In this case, entire cDNAs are fused between Gene III and the appropriate epitope.

**XIX. Identification of PRTS Inhibitors**

Compounds to be tested are arrayed in the wells of a multi-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. PRTS activity is measured for each well and the ability of each compound to inhibit PRTS

activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PRTS activity.

In the alternative, phage display libraries can be used to screen for peptide PRTS inhibitors. Candidates are found among peptides which bind tightly to a protease. In this case, multi-well plate wells are coated with PRTS and incubated with a random peptide phage display library or a cyclic peptide library (Koivunen, E. et al. (1999) Nature Biotech 17:768-774). Unbound phage are washed away and selected phage amplified and rescreened for several more rounds. Candidates are tested for PRTS inhibitory activity using an assay described in Example XVII.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

# Table 1

| Incyte Project ID | Polypeptide SEQ ID NO: | Incyte Polypeptide ID | Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID |
|---|---|---|---|---|
| 5155802 | 1 | 5155802CD1 | 22 | 5155802CB1 |
| 71269782 | 2 | 71269782CD1 | 23 | 71269782CB1 |
| 7472651 | 3 | 7472651CD1 | 24 | 7472651CB1 |
| 7478251 | 4 | 7478251CD1 | 25 | 7478251CB1 |
| 2759385 | 5 | 2759385CD1 | 26 | 2759385CB1 |
| 4226182 | 6 | 4226182CD1 | 27 | 4226182CB1 |
| 5078962 | 7 | 5078962CD1 | 28 | 5078962CB1 |
| 7474340 | 8 | 7474340CD1 | 29 | 7474340CB1 |
| 7477287 | 9 | 7477287CD1 | 30 | 7477287CB1 |
| 2994162 | 10 | 2994162CD1 | 31 | 2994162CB1 |
| 3965293 | 11 | 3965293CD1 | 32 | 3965293CB1 |
| 4948403 | 12 | 4948403CD1 | 33 | 4948403CB1 |
| 7473165 | 13 | 7473165CD1 | 34 | 7473165CB1 |
| 7476667 | 14 | 7476667CD1 | 35 | 7476667CB1 |
| 7479166 | 15 | 7479166CD1 | 36 | 7479166CB1 |
| 3671788 | 16 | 3671788CD1 | 37 | 3671788CB1 |
| 7479181 | 17 | 7479181CD1 | 38 | 7479181CB1 |
| 6621372 | 18 | 6621372CD1 | 39 | 6621372CB1 |
| 4847254 | 19 | 4847254CD1 | 40 | 4847254CB1 |
| 5776350 | 20 | 5776350CD1 | 41 | 5776350CB1 |
| 7473300 | 21 | 7473300CD1 | 42 | 7473300CB1 |

# Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: | Probability score | GenBank Homolog |
|---|---|---|---|---|
| 1 | 5155802CD1 | g7684607 | 0 | [fl][Homo sapiens] calpain 3; calcium activated neutral protease; CAPN3; CL1<br><br>Weilbach,F.X. et al. (1999)Nervenarzt 70:89-100<br>Piechaczyk,M. Methods Mol Biol (2000)144:297-307 |
| 2 | 71269782CD1 | g4539525 | 9.00E-45 | [fl][Homo sapiens] NAALADase II protein<br><br>Pangalos,M.N. et al. (1999) J. Biol. Chem. 274:8470-8483 |
| 3 | 7472651CD1 | g11244759 | 1.00E-144 | [fl][Homo sapiens] ACO protease |
| | | g3649791 | 3.7E-67 | [Homo sapiens] serine protease (TLSP)<br><br>Yoshida,S. et al. (1998)Biochim. Biophys. Acta 1399:225-228 |
| 4 | 7478251CD1 | g3386523 | 1.00E-101 | [fl][Homo sapiens] evolutionarily related interleukin-1beta converting enzyme<br><br>Humke,E.W., Ni,J. and Dixit,V.M. (1998) J. Biol. Chem. 273:15702-15707 |
| 5 | 2759385CD1 | g3220154 | 0 | [5' incom][Homo sapiens] ubiquitin hydrolyzing enzyme I |
| 6 | 4226182CD1 | g1235672 | 1.00E-61 | [fl][Homo sapiens] metalloprotease/disintegrin/cysteine-rich protein precursor<br><br>Weskamp, G. et al. (1996) J. Cell. Biol. 132:717-726 |
| 7 | 5078962CD1 | g6469251 | 9.80E-51 | [Streptomyces coelicolor A3(2)] methionine aminopeptidase (EC |
| 8 | 7474340CD1 | g13429970 | 0 | [fl][Homo sapiens] membrane-type mosaic serine protease |
| | | g6648960 | 1.9E-38 | [Mus musculus] mosaic serine protease epitheliasin<br><br>Jacquinet, E. et al. (2000) FEBS Lett. 468:93-100 |
| 9 | 7477287CD1 | g9798662 | 1.00E-131 | [fl][Suncus murinus] pepsinogen C |
| | | g7008023 | 2.1E-119 | [Callithrix jacchus] pepsinogen C<br><br>Kageyama, T. (2000) J. Biochem. 127:761-770 |

# Table 2 (cont.)

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: | Probability score | GenBank Homolog |
|---|---|---|---|---|
| 10 | 2994162CD1 | g9581879 | 0 | [fl][Homo sapiens] disintegrin metalloproteinase with thrombospondin repeats |
| | | g4929478 | 1.2e-195 | [Rattus norvegicus] a disintegrin and metalloproteinase with thrombospondin motifs 1 |
| 11 | 3965293CD1 | g2739433 | 9.00E-78 | [fl][Mus musculus] hematopoietic-specific IL-2 deubiquitinating enzyme<br><br>Zhu, Y. et al. (1997) J. Biol. Chem. 272:51-57 |
| 12 | 4948403CD1 | g9651704 | 1.00E-168 | [fl][Homo sapiens] carboxypeptidase B precursor |
| | | g203295 | 4.8e-97 | [Rattus norvegicus] carboxypeptidase B |
| 13 | 7473165CD1 | g6467401 | 0 | [Mus musculus] soluble secreted endopeptidase delta<br><br>Ikeda, K. et al. (1999) J. Biol. Chem. 274:32469-32477 |
| 14 | 7476667CD1 | g13560797 | 0 | [fl][Homo sapiens] ubiquitin specific protease |
| | | g2655204 | 2.3e-30 | [Mus musculus] ubiquitin-specific protease |
| 15 | 7479166CD1 | g200507 | 1.70E-60 | [Mus musculus] protease-6<br><br>Serafin, W.E. et al. (1991) J. Biol. Chem. 266:3847-3853 |
| 16 | 3671788CD1 | g10303331 | 0 | [fl][Mus musculus] calpain 12 |
| | | g2570158 | 4.9e-136 | [Mus musculus] m-calpain large subunit<br><br>Muta, T. et al. (1991) J. Biol. Chem. 266:3554-6561 |
| 17 | 7479181CD1 | g217397 | 5.10E-53 | [Tachypleus tridentatus] limulus factor C precursor |
| 18 | 6621372CD1 | g6651071 | 0 | [5' incom][Homo sapiens] disintegrin and metalloproteinase domain 19<br><br>Kurisaki, T. et al. (1998) Mech. Dev. 73:211-215 |
| 19 | 4847254CD1 | g10303329 | 2.00E-76 | [fl][Mus musculus] calpain 12 |
| 20 | 5776350CD1 | g7673618 | 0 | [5' incom][Mus musculus] ubiquitin specific protease |
| 21 | 7473300CD1 | g303704 | 1.00E-06 | [fl][Mus musculus] P100 serine protease of Ra-reactive factor (RaRF) |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 1 | 5155802CD1 | 767 | S154 S320 S322 S329 S352 S375 S384 S496 S511 S527 S552 S557 S590 S642 S655 S90 T13 T291 T361 T574 T713 | N117 N223 N318 N367 N480 N531 | signal_cleavage:M1-A15 | SPSCAN |
| | | | | | CALPAIN CATALYTIC DOMAIN DM01305\|P20807\|19-581: T268-E534, S19-D294 | BLAST_DOMO |
| | | | | | CALPAIN CATALYTIC DOMAIN DM01305\|S57196\|12-574: T268-E534, G21-Y272 | BLAST_DOMO |
| | | | | | CALPAIN CATALYTIC DOMAIN DM01305\|P00789\|3-507: F61-R530 | BLAST_DOMO |
| | | | | | CALPAIN CATALYTIC DOMAIN DM01305\|P07384\|11-517: F61-K529 | BLAST_DOMO |
| | | | | | PROTEASE CALPAIN HYDROLASE SUBUNIT NEUTRAL THIOL LARGE CALCIUM ACTIVATED PROTEINASE CANP PD001545: L74-T369 | BLAST_PRODOM |
| | | | | | PROTEASE CALPAIN HYDROLASE SUBUNIT LARGE NEUTRAL THIOL CALCIUMACTIVATED PROTEINASE CANP PD001874: W381-E534 | BLAST_PRODOM |
| | | | | | CALPAIN SUBUNIT PROTEASE NEUTRAL CALCIUM ACTIVATED PROTEINASE BINDING CALCIUM LARGE PD002827: L666-I729 | BLAST_PRODOM |
| | | | | | CANP HYDROLASE LARGE PD003609: E595-F663 | BLAST_PRODOM |
| | | | | | CALPAIN SUBUNIT CALCIUM BINDING PROTEASE CALCIUM ACTIVATED PROTEINASE CANP HYDROLASE LARGE | BLAST_PRODOM |
| | | | | | EF-hand calcium-binding domain protein BL00018: D651-F663 | BLIMPS_BLOCKS |
| | | | | | Calpain cysteine protease (C2) family signature PR00704: K59-P82, W99-I121, Q123-T139, Y159-T184, L189-L212, G214-I241, E345-C366, S395-Y412, R500-E528 | BLIMPS_PRINTS |

# Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | Calpain family cysteine protease Peptidase_C2: L74-T369 | HMMER_PFAM |
| | | | | | EF hand: S642-I670, A672-A700 | HMMER_PFAM |
| | | | | | Calpain large subunit, domain III Calpain_III: T380-E534, | HMMER_PFAM |
| | | | | | EF-Hand calcium binding domain: D651-F663, D681-M693 | MOTIFS |
| | | | | | Eukaryotic Thiol (cysteine) Proteases Active site: Q123-A134 | MOTIFS |
| 2 | 71269782CD1 | 574 | S117 S180 S197 S255 S267 S315 S362 S366 S393 S404 S59 S92 T271 T398 T44 T440 Y106 | N10 N216 N295 N373 N534 | PROTEIN AMINOPEPTIDASE ANTIGEN RECEPTOR TRANSMEMBRANE MEMBRANE CARBOXYPEPTIDASE TRANSFERRIN HYDROLASE PROSTATE SPECIFIC PD001808: N410-T556, K179-S218 | BLAST_PRODOM |
| | | | | | transmembrane domain: I128-V146 | HMMER |
| 3 | 7472651CD1 | 320 | S166 S211 S220 S226 S288 T153 T242 T297 T316 | N235 N296 | trypsin: L86-I313 | HMMER_PFAM |
| | | | | | Serine proteases, trypsin family active sites: Trypsin_Histidine: L122-C127 | MOTIFS |
| | | | | | TRYPSIN DM00018 \| P12788 \| 23-243: K85-M317 | BLAST_DOMO |
| | | | | | TRYPSIN DM00018 \| P00764 \| 8-225: L86-M317 | BLAST_DOMO |
| | | | | | TRYPSIN DM00018 \| P35031 \| 20-238: K85-M317 | BLAST_DOMO |
| | | | | | TRYPSIN DM00018 \| S49489 \| 21-238: L86-M317 | BLAST_DOMO |
| | | | | | PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046:R133-I313, L86-Y248 | BLAST_PRODOM |
| | | | | | Serine proteases, trypsin family, histidine proteins BL00134: C111-C127, E267-G290, P300-I313 | BLIMPS_BLOCKS |
| | | | | | Type I fibronectin domain proteins BL01253: C111-A124, A266-C279 | BLIMPS_BLOCKS |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | Kringle domain proteins. BL00021: C111-Q128 | BLIMPS_BLOCKS |
| | | | | | Chymotrypsin serine protease family (S1) signature PR00722: G112-C127, S166-A180, A266-V278 | BLIMPS_PRINTS |
| | | | | | Serine proteases, trypsin family, active sites for: Trypsin_Histidine: L103-P147 Trypsin-Serine: L252-D295 | PROFILESCAN |
| 4 | 7478251CD1 | 378 | S102 S154 S244 S271 S313 S52 S79 T118 T134 T179 T20 T232 Y125 Y147 Y170 | N152 N177 N311 N319 | Caspase recruitment domain CARD: A2-S91 | HMMER_PFAM |
| | | | | | ICE-like protease (caspase) p20 domain ICE_p20: K131-I264 | HMMER_PFAM |
| | | | | | ICE-like protease (caspase) p10 domain ICE_p10:A291-P376, | HMMER_PFAM |
| | | | | | INTERLEUKIN-1 BETA CONVERTING ENZYME FAMILY HISTIDINE DM01067\|P49662\|97-280: Q97-W266 | BLAST_DOMO |
| | | | | | INTERLEUKIN-1 BETA CONVERTING ENZYME FAMILY HISTIDINE DM01067\|B57511\|138-321: Q97-W266 | BLAST_DOMO |
| | | | | | INTERLEUKIN-1 BETA CONVERTING ENZYME FAMILY HISTIDINE DM01067\|P51878\|138-321: Q97-W266 | BLAST_DOMO |
| | | | | | INTERLEUKIN-1 BETA CONVERTING ENZYME FAMILY HISTIDINE DM01067\|P29466\|124-307: G103-G275 | BLAST_DOMO |
| | | | | | PRECURSOR PROTEASE HYDROLASE THIOL ZYMOGEN APOPTOSIS PROTEIN APOPTOTIC CASPASE1 CYSTEINE PD001408: K131-N260 | BLAST_PRODOM |
| | | | | | CASPASE12 PRECURSOR EC 3.4.22. HYDROLASE THIOL PROTEASE APOPTOSIS ZYMOGEN PD103766:V11-K131 | BLAST_PRODOM |
| | | | | | Caspase family histidine proteins BL01121: L148-M183, E195-S210, C242-G259, K294-I328, L340-E352 | BLIMPS_BLOCKS |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | INTERLEUKIN-1B CONVERTING ENZYME SIGNATURE PR00376: R133-N146, R151-G169, G169-L187, T202-S210, C242-N260, S313-I324, L366-F375 | BLIMPS_PRINTS |
| | | | | | Caspase family active site: Ice_Serine: K248-G259 | MOTIFS |
| 5 | 2759385CD1 | 366 | S17 S189 S190 S216 S234 S271 T131 T2 T285 T89 Y358 | N15 N178 N205 N284 | Ubiquitin carboxyl-terminal hydrolase family 1 UCH-1: F35-Y66 | HMMER_PFAM |
| | | | | | Ubiquitin carboxyl-terminal hydrolase family 2 UCH-2: L292-S364 | HMMER_PFAM |
| | | | | | UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2 DM00659\|P39967\|359-610: K72-G306 | BLAST_DOMO |
| | | | | | UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2 DM00659\|P40818\|782-1103: G41-E341 | BLAST_DOMO |
| | | | | | PROTEASE UBIQUITIN HYDROLASE ENZYME UBIQUITINSPECIFIC CARBOXYLTERMINAL DEUBIQUITINATING THIOLESTERASE PROCESSING CONJUGATION PD000590: G36-S189 | BLAST_PRODOM |
| | | | | | PROTEASE UBIQUITIN HYDROLASE UBIQUITINSPECIFIC ENZYME DEUBIQUITINATING CARBOXYLTERMINAL THIOLESTERASE PROCESSING CONJUGATION PD017412: S190-L282 | BLAST_PRODOM |
| | | | | | Ubiquitin carboxyl-terminal hydrolase family 2 BL00972: G36-L53, Y116-L125, V168-C182, Y296-S320, H321-E342 | BLIMPS_BLOCKS |
| | | | | | Ubiquitin carboxyl-terminal hydrolase family 2 signatures Uch_2_1: G36-Q51 Uch_2_2: Y296-Y314 | MOTIFS |

# Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 6 | 4226182CD1 | 389 | S138 S140 S215 S285 S291 S32 S337 S350 S369 S61 S82 S97 T173 T204 T363 T373 | N213 N80 | Disintegrin signature disintegrin: A22-C86 | HMMER_PFAM |
| | | | | | do ZINC; REGULATED; EPIDIDYMAL; NEUTRAL; DM00591\|S47656\|462-624: C79-A210 | BLAST_DOMO |
| | | | | | TRANSMEMBRANE METALLOPROTEASE SIGNAL PRECURSOR PROTEIN GLYCOPROTEIN CELL FERTILIN BETA ADHESION PD001269: N94-I163 | BLAST_PRODOM |
| | | | | | CELL ADHESION PLATELET BLOOD COAGULATION VENOM DISINTEGRIN METALLOPROTEASE PRECURSOR SIGNAL PD000664: C28-C86 | BLAST_PRODOM |
| | | | | | DISINTEGRIN SIGNATURE PR00289: C47-R66, E77-D89 | BLIMPS_PRINTS |
| | | | | | transmembrane domain: W298-A318 | HMMER |
| | | | | | Disintegrins signature disintegrins.prf: G8-D89 | PROFILESCAN |
| 7 | 5078962CD1 | 217 | T2 T203 | N151 | metallopeptidase family M24 Peptidase_M24: M1-Q208 | HMMER_PFAM |
| | | | | | AMINOPEPTIDASE HYDROLASE METHIONINE PEPTIDASE PROTEIN COBALT M DIPEPTIDASE XPRO MAP PD000555: E4-D181 | BLAST_PRODOM |
| | | | | | Aminopeptidase P and proline dipeptidase proteins BL00491C: M157-E171 | BLIMPS_BLOCKS |
| | | | | | Methionine aminopeptidase subfamily 1 BL00680: D55-F76 | BLIMPS_BLOCKS |
| | | | | | METHIONINE AMINOPEPTIDASE-1 SIGNATURE PR00599: V33-P46, D55-D71, F125-G137, L155-P167 | BLIMPS_PRINTS |
| | | | | | METHIONINE AMINOPEPTIDASE DM01530\|Q01662\|123-375: M1-T211 | BLAST_DOMO |
| | | | | | Methionine aminopeptidase signature map.prf: I112-I168 | PROFILESCAN |

# Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 8 | 7474340CD1 | 486 | S101 S252 S254 S301 S391 S96 T153 T289 T318 T349 T402 T428 T465 | N250 N287 N400 | Trypsin family active site trypsin: I321-H438 | HMMER_PFAM |
| | | | | | TRYPSIN DM00018\|P26262\|391-624: I321-P429 | BLAST_DOMO |
| | | | | | TRANSMEMBRANE PROTEASE, SERINE 2 EC 3.4.21. HYDROLASE PROTEASE SIGNALANCHOR PD072395: P86-R320 | BLAST_PRODOM |
| | | | | | PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: I321-S463 | BLAST_PRODOM |
| | | | | | Serine proteases, trypsin family BL00134A: C346-C362 | BLIMPS_BLOCKS |
| | | | | | Kringle domain proteins BL00021B: C346-F363 | BLIMPS_BLOCKS |
| | | | | | CHYMOTRYPSIN SERINE PROTEASE PR00722: E405-L419, G347-C362 | BLIMPS_PRINTS |
| | | | | | transmembrane domain: L163-W184 | HMMER |
| | | | | | Trypsin family active site Trypsin_His: L357-C362 | MOTIFS |
| | | | | | Serine proteases, trypsin family, active sites trypsin_his.prf: W334-A389 | PROFILESCAN |
| 9 | 7477287CD1 | 390 | S164 S175 S27 S375 T123 | N311 | Eukaryotic aspartyl protease asp: P65-V89, P101-S389 | HMMER_PFAM |
| | | | | | EUKARYOTIC AND VIRAL ASPARTYL PROTEASES DM00126\|P20142\|17-386: R19-A387 | BLAST_DOMO |
| | | | | | PROTEASE ASPARTYL HYDROLASE PRECURSOR SIGNAL ZYMOGEN GLYCOPROTEIN ASPARTIC PROTEINASE MULTIGENE PD000182: P65-A387 | BLAST_PRODOM |
| | | | | | Eukaryotic and viral aspartyl protease BL00141: D178-A189, G230-G239, I364-A387 | BLIMPS_BLOCKS |
| | | | | | PEPSIN (A1) ASPARTIC PROTEASE PR00792: T80-L100, G225-s238, A275-V286, 2363-D378 | BLIMPS_PRINTS |
| | | | | | Signal_peptide | HMMER |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 10 | 2994162CD1 | 1916 | S122 S171 S27 S400 S460 S59 S732 S781 S782 S811 S924 S947 S968 T139 T156 T199 T220 T25 T262 T266 T344 T370 T391 T53 T545 T758 T771 T815 T823 T893 T914 T953 T998 T1155 T1159 T1008 T1019 S1122 S1189 S1196 S1257 T1267 S1329 T1343 S1393 S1455 T1509 T1522 S1526 T1539 T1551 S1579 S1619 T1625 T1661 T1687 T1707 S1789 T1840 S1865 S1869 T1909 Y164 Y1263 Y1521 Y1558 Y1634 Y1735 | N116 N252 N730 N821 N93 N1194 N1248 N1769 N1787 | Reprolysin (M12B) family zinc metalloprotease Reprolysin: R274-P480 | HMMER_PFAM |
| | | | | | Reprolysin family propeptide Pep_M12B_propep: N93-R223 | HMMER_PFAM |
| | | | | | Thrombospondin type 1 domain tsp_1: G570-C623, W1313-C1364, W1426-C1479 | HMMER_PFAM |
| | | | | | Neutral zinc metallopeptidase BL00142: T412-N422 | BLIMPS_BLOCKS |
| | | | | | do ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN; DM00368\|S60257\|204-414: L270-E481 | BLAST_DOMO |
| | | | | | METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION PD000791: L270-P480 | BLAST_PRODOM |
| | | | | | PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD014161: K734-I851 PD011654: I661-C733 | BLAST_PRODOM |
| | | | | | Zinc_Protease: T412-F421 | MOTIFS |
| 11 | 3965293CD1 | 314 | S22 S23 S272 S284 S294 S311 S36 S71 S72 T47 | N92 | Ubiquitin carboxyl-terminal hydrolases family 1 UCH-1: A80-R111 | HMMER_PFAM |
| | | | | | Ubiquitin carboxyl-terminal hydrolases family 2 BL00972: G81-L98, G156-L165, I193-C207 | BLIMPS_BLOCKS |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2<br>DM00659\|P50102\|141-420: Q158-F283<br>DM00659\|Q09738\|149-388: N84-F283 | BLAST_DOMO |
| | | | | | PROTEASE UBIQUITIN HYDROLASE ENZYME UBIQUITINSPECIFIC CARBOXYLTERMINAL DEUBIQUITINATING THIOLESTERASE PROCESSING CONJUGATION<br>PD000590:L62-H120, F153-T216<br>PD017412: F217-F283 | BLAST_PRODOM |
| | | | | | Ubiquitin carboxyl-terminal hydrolases family 2 Uch_2_1: G81-Q96 | MOTIFS |
| 12 | 4948403CD1 | 437 | S141 S299 S335 S381 S60 T124 T216 T417 T49 T80 Y352 Y54 | N153 N427 N89 | Zinc carboxypeptidase Zn_carbOpept: Y139-E420 | HMMER_PFAM |
| | | | | | Zinc carboxypeptidases, zinc-binding regions BL00132: Y139-L179, R187-W200, Y217-R257, S261-K275, P287-H313, H316-K337, T373-G390 | BLIMPS_BLOCKS |
| | | | | | ZINC CARBOXYPEPTIDASES, ZINC-BINDING REGION 1 DM00683\|P19223\|107-414: S132-L432 | BLAST_DOMO |
| | | | | | CARBOXYPEPTIDASE PRECURSOR SIGNAL HYDROLASE ZINC ZYMOGEN PROTEIN GP180CARBOXYPEPTIDASE PD001916: Y139-F344, G194-C418<br>PD005637: H31-K118 | BLAST_PRODOM |
| | | | | | CARBOXYPEPTIDASE A METALLOPROTEASE FAMILY SIGNATURE PR00765: I165-L177, R187-I201, G267-K275, L321-Y334 | BLIMPS_PRINTS |
| | | | | | Carboxypept_Zn_2: H324-Y334 | MOTIFS |
| | | | | | Zinc carboxypeptidases, zinc-binding regions signatures carboxypept_zn_2.prf: E302-L358 | PROFILESCAN |
| | | | | | signal_cleavage: M1-S30 | SPSCAN |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 13 | 7473165CD1 | 742 | S102 S144 S151 S209 S234 S326 S356 S377 S410 S431 S457 S467 S515 S689 S698 T123 T394 T446 T636 Y407 Y490 Y616 | N121 N142 N172 N208 N315 N494 N601 N620 | Peptidase family M13 Peptidase_M13: N535-V741 | HMMER_PFAM |
| | | | | | Neutral zinc metallopeptidases BL00142: V573-D583 | BLIMPS_BLOCKS |
| | | | | | NEPRILYSIN DM02569\|P08473\|11-748: L20-W742 | BLAST_DOMO |
| | | | | | PROTEIN ZINC METALLOPROTEASE HYDROLASE TRANSMEMBRANE GLYCOPROTEIN SIGNALANCHOR ENDOPEPTIDASE NEUTRAL ENZYME PD001606: E240-P692 PD002031: A62-P245 | BLAST_PRODOM |
| | | | | | NEPRILYSIN METALLOPROTEASES PR00786: L527-S539, I545-F557, N566-F582, E639-A650 | BLIMPS_PRINTS |
| | | | | | zinc_Protease: V573-F582 | MOTIFS |
| | | | | | transmem_domain: L20-Y38 | HMMER |
| | | | | | signal_cleavage:M1-V32 | SPSCAN |
| 14 | 7476667CD1 | 582 | S203 S222 S273 S328 S350 S357 S358 S367 S376 S400 S432 S44 S470 S474 S523 S565 S566 S71 T134 T188 T221 T244 T29 T438 T6 T91 | N32 N468 N520 | Ubiquitin carboxyl-terminal hydrolase family 2 UCH-2: I484-Q544 | HMMER_PFAM |
| | | | | | Ubiquitin carboxyl-terminal hydrolase family BL00972: I487-N511, N513-T534 | BLIMPS_BLOCKS |
| | | | | | do UBIQUITIN; TRANSFORMING; HYDROLASE; TERMINAL; DM08764\|P35125\|548-820: L45-R318 | BLAST_DOMO |
| | | | | | UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2 DM00659\|P40818\|782-1103: A206-D294, Y488-L540 DM00521\|P35125\|1007-1051: L500-Q545 | BLAST_DOMO |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | UBIQUITIN CARBOXYLTERMINAL HYDROLASE 6 THIOLESTERASE UBIQUITINSPECIFIC PROCESSING PROTEASE DEUBIQUITINATING ENZYME PROTOONCOGENE TRE2 CONJUGATION THIOL MULTIGENE FAMILY PD085597: R378-I487 PD038816: I55-S203 PD119604: M1-I54 PD085589: C524-Q582 | BLAST_PRODOM |
| | | | | | Uch_2_2 Y488-Y505 | MOTIFS |
| 15 | 7479166CD1 | 290 | S250 S54 S91 T264 Y133 | N150 N209 | Trypsin active sites trypsin: I75-S177, P186-I282 | HMMER_PFAM |
| | | | | | Serine proteases, trypsin family BL00134: C106-C122, D233-V256, P269-I282 | BLIMPS_BLOCKS |
| | | | | | Type I fibronectin domain BL01253: C106-A119, R232-C245, W251-Q285 | BLIMPS_BLOCKS |
| | | | | | Kringle domain proteins BL00021: C106-I123, G241-I282 | BLIMPS_BLOCKS |
| | | | | | CHYMOTRYPSIN SERINE PROTEASES PR00722: G107-C122, G164-P178, R232-V244 | BLIMPS_PRINTS |
| | | | | | TRYPSIN DM00018\|P21845\|31-271: G74-P186, E182-V286 | BLAST_DOMO |
| | | | | | PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: P187-I282, I75-S180 | BLAST_PRODOM |
| | | | | | Trypsin family active sites: Trypsin_His: L117-C122 Trypsin_Ser: D233-V244 | MOTIFS |
| | | | | | Serine proteases, trypsin family, active sites trypsin_his.prf: A103-G147 trypsin_ser.prf: I220-L265 | PROFILESCAN |
| | | | | | signal_cleavage: M1-A60 | SPSCAN |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 16 | 3671788CD1 | 708 | S244 S488 S5 S67 S93 T266 T388 T421 T459 T461 T492 T577 T654 T697 | N556 | Calpain family cysteine protease Peptidase_C2: L45-S341 | HMMER_PFAM |
| | | | | | Calpain large subunit, domain III Calpain_III: G353-A499 | HMMER_PFAM |
| | | | | | CALPAIN CYSTEINE PROTEASE PR00704: Q30-A53, W75-V97, Q99-T115, Y135-V160, L165-L188, G190-L217, E317-C338, N368-F385 | BLIMPS_PRINTS |
| | | | | | PROTEASE CALPAIN HYDROLASE SUBUNIT NEUTRAL THIOL LARGE CALCIUMACTIVATED PROTEINASE PD001545: L45-S341 PD002827: L607-V670 PD001874: W354-E401, C424-Y491 | BLAST_PRODOM |
| | | | | | CALPAIN CATALYTIC DOMAIN DM01305\|P17655\|1-505: D14-G402, C424-N463 | BLAST_DOMO |
| | | | | | CALPAIN CATALYTIC DOMAIN DM01305\|A48764\|1-507: M1-G402, G418-Q454 | BLAST_DOMO |
| | | | | | Cysteine protease active site Thiol_Protease_Cys: Q99-A110 | MOTIFS |
| | | | | | EF hand calcium binding domain Ef_Hand: D622-L634 | MOTIFS |
| | | | | | Trypsin active site trypsin: W391-I644 | HMMER_PFAM |
| 17 | 7479181CD1 | 649 | S257 S353 S354 S365 S402 S502 S519 S552 S571 S627 S93 T102 T318 T361 T545 T86 | N380 N543 N96 | CUB domain CUB: C128-Y233 | HMMER_PFAM |
| | | | | | EGF-like domain EGF: C239-C271 | HMMER_PFAM |
| | | | | | Serine proteases, trypsin family, active sites trypsin_his_prf: K411-E464 | PROFILESCAN |
| | | | | | Serine proteases, trypsin family (p<0.0012) BL00134: C418-C434, S631-I644 | BLIMPS_BLOCKS |
| | | | | | CUB domain proteins BL01180B: C177-G187 (p<0.13) | BLIMPS_BLOCKS |
| | | | | | Kringle domain proteins BL00021B: C418-V435 (p<0.087) | BLIMPS_BLOCKS |

# Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | Type II EGF-like signature PR00010: E235-H246, G256-Y266, T267-N273 | BLIMPS_PRINTS |
| | | | | | CHYMOTRYPSIN SERINE PROTEASE PR00722: S419-C434, L485-A499 | BLIMPS_PRINTS |
| | | | | | Sushi domain proteins (Short consensus repeat) PF00084: H336-F347, G362-C371 | BLIMPS_PFAM |
| | | | | | TRYPSIN DM00018\|P28175\|759-1018: R390-R646 | BLAST_DOMO |
| | | | | | PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: W391-I644 | BLAST_PRODOM |
| | | | | | signal_peptide: M1-A32 | HMMER |
| | | | | | EGF-like domain Egf: C260-C271 | MOTIFS |
| 18 | 6621372CD1 | 918 | S208 S284 S364 S38 S647 S787 S823 S830 S831 S90 S907 S915 T105 T106 T118 T131 T182 T194 T449 T488 T504 T520 T764 Y115 Y243 | N144 N444 N447 N645 | Reprolysin (M12B) family zinc metallopeptidase Reprolysin: K210-P408 | HMMER_PFAM |
| | | | | | Reprolysin family propeptide Pep_M12B_propep: D79-K195 | HMMER_PFAM |
| | | | | | Disintegrin signature disintegrin: E425-Q500 | HMMER_PFAM |
| | | | | | Disintegrin signature disintegrins.prf: E436-P495 | PROFILESCAN |
| | | | | | Neutral zinc metallopeptidases, zinc-binding region signature zinc_protease.prf: S325-G377 | PROFILESCAN |
| | | | | | Neutral zinc metallopeptidases BL00142: T342-G352 | BLIMPS_BLOCKS |
| | | | | | DISINTEGRIN SIGNATURE PR00289: C456-R475, E485-N497 | BLIMPS_PRINTS |
| | | | | | NEPRILYSIN METALLOPROTEASE PR00786C: N335-F351 | BLIMPS_PRINTS |
| | | | | | MELTRIN, BETA METALLOPROTEASEDISINTEGRIN MELTRIN BETA INTEGRIN PROTEASE METALLOPROTEASE PD105322: P696-G888 PD171676: K571-C643 | BLAST_PRODOM |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION PD000791: K210-P408 | BLAST_PRODOM |
| | | | | | CELL ADHESION PLATELET BLOOD COAGULATION VENOM DISINTEGRIN METALLOPROTEASE PRECURSOR SIGNAL PD000664: E425-Y499 | BLAST_PRODOM |
| | | | | | do ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN; DM00368\|S60257\|204-414: K202-D409 | BLAST_DOMO |
| | | | | | do ZINC; REGULATED; EPIDIDYMAL; NEUTRAL; DM00591\|S60257\|492-628: F486-L625 | BLAST_DOMO |
| | | | | | Zinc_Protease: T342-F351 | MOTIFS |
| | | | | | transmembrane domain: V700-Y721 | HMMER |
| | | | | | signal_cleavage: M1-P22 | SPSCAN |
| 19 | 4847254CD1 | 218 | T164 T207 T49 | N28 | CALPAIN CATALYTIC DOMAIN DM01221\|P20807\|719-819: L117-F217 DM01221\|S57196\|708-808: L117-F217 DM01221\|P00789\|602-702: L117-M212 DM01221\|P07384\|612-712: L117-F217 | BLAST_DOMO |
| | | | | | CALPAIN SUBUNIT PROTEASE NEUTRAL CALCIUMBINDING CALCIUMACTIVATED PROTEINASE CANP HYDROLASE LARGE PD002827: L117-V180 | BLAST_PRODOM |
| | | | | | Calcium binding domain E£_Hand: D132-L144 | MOTIFS |
| | | | | | Calcium binding domain efhand: E123-A151 | HMMER_PFAM |
| | | | | | signal_cleavage: M1-T47 | SPSCAN |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 20 | 5776350CD1 | 656 | S141 S145 S22 S272 S279 S301 S338 S410 S483 S493 S510 S520 S524 S572 S624 S95 S99 T107 T171 T204 T260 T451 T502 T529 T634 Y640 | N16 | Ubiquitin carboxyl-terminal hydrolase family 1<br>UCH-1: R308-D339<br>Ubiquitin carboxyl-terminal hydrolase family 2<br>UCH-2: N590-K650<br>Ubiquitin carboxyl-terminal hydrolase family 2<br>BL00972: G309-L326, Y390-L399, I429-C443, K593-Q617, K619-Y640<br>UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2<br>DM00659\|P40818\|782-1103: S493-L646, L313-N421, I428-L463<br>PROTEASE UBIQUITIN HYDROLASE UBIQUITINSPECIFIC ENZYME DEUBIQUITINATING CARBOXYLTERMINAL THIOLESTERASE PROCESSING CONJUGATION<br>PD017412: S493-P583<br>Ubiquitin carboxyl-terminal hydrolase family 1 Uch_2_1: G309-Q324<br>Ubiquitin carboxyl-terminal hydrolase family 2 Uch_2_2: Y594-Y611 | HMMER_PFAM<br>HMMER_PFAM<br>BLIMPS_BLOCKS<br>BLAST_DOMO<br>BLAST_PRODOM<br>MOTIFS<br>MOTIFS<br>HMMER_PFAM |
| 21 | 7473300CD1 | 509 | S137 S156 S488 T130 T163 T32 T37 T41 Y286 | N253 N33 N394 | Trypsin family serine protease active site<br>trypsin: K279-P358<br>Trypsin family serine protease active site<br>trypsin_his.prf: I297-P343<br>Trypsin family serine protease active site<br>BL00134A: C305-C321<br>Kringle domain proteins BL00021B: C305-V322 | PROFILESCAN<br>BLIMPS_BLOCKS<br>BLIMPS_BLOCKS |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | CHYMOTRYPSIN SERINE PROTEASE ACTIVE SITE PR00722A: S306-C321 | BLIMPS_PRINTS |
| | | | | | Trypsin family serine protease active site Trypsin_His: L316-C321 | MOTIFS |

# Table 4

| Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID | Sequence Length | Selected Fragment(s) | Sequence Fragments | 5' Position | 3' Position |
|---|---|---|---|---|---|---|
| 22 | 5155802CB1 | 2789 | 1-1939 | 71666762V1 | 1728 | 2444 |
| | | | | 71668725V1 | 1024 | 1733 |
| | | | | 8001825H1 (LNODTUC02) | 1 | 383 |
| | | | | 71668385V1 | 248 | 960 |
| | | | | 8089190H1 (BRACDIK08) | 2133 | 2789 |
| | | | | 71667190V1 | 928 | 1658 |
| | | | | 70239197V1 | 1712 | 2237 |
| | | | | 70235564V1 | 472 | 1009 |
| 23 | 71269782CB1 | 2267 | 1701-2267 | 70900108V1 | 586 | 1167 |
| | | | | 70899845V1 | 1716 | 2234 |
| | | | | 71269782V1 | 1286 | 1963 |
| | | | | GBI.g8567524_edit | 1142 | 2267 |
| | | | | 2779031F6 (OVARTUT03) | 10 | 573 |
| | | | | g7377067 | 1 | 396 |
| | | | | 70899669V1 | 360 | 977 |
| | | | | 71874795V1 | 1142 | 1708 |
| 24 | 7472651CB1 | 963 | 720-801, 1-665, 838-912 | FL7472651_g7689999_000022_g3649791 | 1 | 963 |
| 25 | 7478251CB1 | 1137 | 1-489, 779-876 | 72001656V1 | 779 | 1137 |
| | | | | g8117619_edit_1 | 1 | 80 |
| | | | | g8117619_edit_2 | 256 | 778 |
| | | | | 72004235V1 | 3 | 261 |
| 26 | 2759385CB1 | 3204 | 2123-2558, 1-72, 505-529, 3127-3204 | 6983266H1 (BRAIFER05) | 845 | 1382 |
| | | | | 1275720T6 (TESTTUT02) | 2453 | 3117 |
| | | | | 2759385F6 (THP1AZS08) | 1329 | 1748 |
| | | | | 7168141H1 (MCLRNOC01) | 413 | 929 |
| | | | | 3690313F6 (HEAANOT01) | 1 | 475 |
| | | | | 2732484H1 (OVARTUT04) | 2934 | 3126 |
| | | | | 7380327H1 (ENDMUNE01) | 1570 | 2128 |
| | | | | 647852H1 (CARCTXT02) | 686 | 948 |
| | | | | 4520886H1 (SINJNOT03) | 2950 | 3204 |
| | | | | 659258R6 (BRAINOT03) | 2006 | 2493 |
| | | | | 6263739H1 (MCLDTXN03) | 2220 | 2556 |
| | | | | 2759385R6 (THP1AZS08) | 949 | 1563 |

Table 4 (cont.)

| Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID | Sequence Length | Selected Fragment(s) | Sequence Fragments | 5' Position | 3' Position |
|---|---|---|---|---|---|---|
| 27 | 4226182CB1 | 1641 | 1-696 | 645682T6 (BRSTTUT02) | 984 | 1631 |
| | | | | 5015693F6 (BRAXNOT03) | 372 | 1008 |
| | | | | 55062402J1 | 1 | 545 |
| | | | | 71975126V1 | 655 | 1054 |
| | | | | 645682F1 (BRSTTUT02) | 1068 | 1641 |
| 28 | 5078962CB1 | 1983 | 1-319, 1809-1983 | 2937276F6 (THYMFET02) | 780 | 1385 |
| | | | | 55058283J2 | 1 | 761 |
| | | | | 6473257H1 (PLACFEB01) | 1050 | 1723 |
| | | | | 8118369H1 (TONSDIC01) | 686 | 1341 |
| | | | | 6508675H1 (BRAHNOT02) | 1477 | 1983 |
| 29 | 7474340CB1 | 1574 | 1-37 | 5558974T9 (TONSDIT01) | 829 | 1350 |
| | | | | 55068051J1 | 426 | 1098 |
| | | | | g2056077 | 1134 | 1574 |
| | | | | 55068054J1 | 1 | 602 |
| 30 | 7477287CB1 | 1173 | 1-732, 1112-1173, 834-1071 | g8546678_edit_01 | 1 | 100 |
| | | | | g8546678_edit_02 | 225 | 1173 |
| | | | | 825016H1_edit_1 (PROSNOT06) | 55 | 224 |
| 31 | 2994162CB1 | 6013 | 5667-6013, 2770-4197, 683-2187, 1-103, 219-247 | 3071581H1 (UTRSNOR01) | 3391 | 3614 |
| | | | | 7122715H1 (BRAHNOE01) | 2267 | 2792 |
| | | | | 71229995V1 | 5281 | 5880 |
| | | | | 7992663H1 (UTRSDIC01) | 4366 | 5039 |
| | | | | 6177981F6 (BMARUNT02) | 145 | 777 |
| | | | | 70867656V1 | 5401 | 6013 |
| | | | | 6706152H1 (HEAADIR01) | 4733 | 5393 |
| | | | | 496053H1 (HNT2NOT01) | 2881 | 3243 |
| | | | | g7242978_CD | 433 | 4914 |
| | | | | 5301201H1 (MUSCNOT11) | 2035 | 2300 |
| | | | | 7407622H1 (UTREDME05) | 405 | 952 |
| | | | | 7606552H1 (COLRTUE01) | 3834 | 4394 |
| | | | | 7272409H1 (OVARDIJ01) | 3571 | 4162 |
| | | | | 7090903F6 (BRAUTDR03) | 1153 | 1733 |
| | | | | 7100145R6 (BRAWTDR02) | 1248 | 2171 |
| | | | | 55062765H1 | 1 | 245 |
| | | | | 7100145F6 (BRAWTDR02) | 798 | 1652 |
| | | | | 7728093J1 (UTRCDIE01) | 2412 | 3040 |

Table 4 (cont.)

| Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID | Sequence Length | Selected Fragment(s) | Sequence Fragments | 5' Position | 3' Position |
|---|---|---|---|---|---|---|
| 32 | 3965293CB1 | 1393 | 397-1002 | 3965293F6 (PROSNOT14) | 1 | 858 |
| | | | | 71832720V1 | 651 | 1393 |
| 33 | 4948403CB1 | 1993 | 1654-1687, 1-123, 850-1300 | 4600759H1 (COLSTUT01) | 1025 | 1282 |
| | | | | 71982269V1 | 1420 | 1993 |
| | | | | 5763587T7 (PROSBPT02) | 657 | 1179 |
| | | | | 70484250V1 | 1180 | 1790 |
| | | | | GBI.g8080699_000017_00 0013.edit | 528 | 974 |
| | | | | 5763587F7 (PROSBPT02) | 1 | 473 |
| | | | | 7930210H1 (COLNDIS02) | 116 | 619 |
| 34 | 7473165CB1 | 2318 | 1-1362, 1756-2138 | 2250635H1 (OVARTUT01) | 2193 | 2318 |
| | | | | GBI.g9367391_000005_00 0006.edit | 1848 | 2318 |
| | | | | FL7473165-g7329540_000015-g6467401 | 1020 | 1259 |
| | | | | 55072914H1 | 272 | 891 |
| | | | | 55073757J1 | 1 | 465 |
| | | | | 55062846H1 | 452 | 1124 |
| | | | | GBI:g8039388_000002.edit | 1161 | 1982 |
| 35 | 7476667CB1 | 1931 | 1909-1931 | 337733R6 (EOSIHET02) | 1418 | 1931 |
| | | | | 160823 4T6 (LUNGNOT15) | 1301 | 1930 |
| | | | | 71729901V1 | 678 | 1385 |
| | | | | 71734439V1 | 608 | 1345 |
| | | | | 55027506H1 (ADMEDNV30) | 1 | 687 |
| 36 | 7479166CB1 | 1218 | 1-299, 369-666, 1020-1057, 739-762 | g4394411 | 764 | 1218 |
| | | | | GNN.g7635593_000002_00 6 | 1 | 873 |
| 37 | 3671788CB1 | 2679 | 1-1760 | 72038124V1 | 1950 | 2679 |
| | | | | 6198936H1 (PITUNON01) | 1721 | 2372 |
| | | | | 3671788T7 (KIDNTUT16) | 348 | 864 |
| | | | | 6431661H1 (LUNGNON07) | 1792 | 2390 |
| | | | | 526464H1 (EOSINOT02) | 1680 | 1777 |

# Table 4 (cont.)

| Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID | Sequence Length | Selected Fragment(s) | Sequence Fragments | 5' Position | 3' Position |
|---|---|---|---|---|---|---|
| 37 | | | | GBI.g8576128_000022_000025.edit | 131 | 2257 |
| | | | | 2579533T6 (KIDNTUT13) | 1 | 439 |
| | | | | 7729129H1 (UTRCDIE01) | 586 | 1196 |
| 38 | 7479181CB1 | 2632 | 1-1603 | 1681388F7 (STOMFET01) | 2423 | 2632 |
| | | | | 8113752H1 (OSTEUNC01) | 1 | 515 |
| | | | | 71510880V1 | 1282 | 2009 |
| | | | | 70737244V1 | 448 | 1028 |
| | | | | 71509933V1 | 1892 | 2626 |
| | | | | 7245927H1 (PROSTMY01) | 2037 | 2628 |
| | | | | 70733946V1 | 575 | 1238 |
| | | | | 71511332V1 | 1216 | 1920 |
| 39 | 6621372CB1 | 2757 | 2517-2757, 430-1288 | 7715927J1 (SINTFEE02) | 781 | 1531 |
| | | | | 5456122H1 (SINITUT03) | 2606 | 2757 |
| | | | | 6887315F6 (BRAITDR03) | 1700 | 2324 |
| | | | | 7372052H2 (BRAIFEE04) | 2235 | 2722 |
| | | | | g6651070_CD | 293 | 2705 |
| | | | | GBI.g7709272_g6651070_g7709257_edit | 1 | 2757 |
| | | | | 7723192J2 (THYRDIE01) | 1096 | 1691 |
| | | | | 8037549H1 (SMCRUNE01) | 397 | 1010 |
| | | | | 8037549J1 (SMCRUNE01) | 1647 | 2311 |
| 40 | 4847254CB1 | 1892 | 1-764, 1773-1892, 918-1029 | 4847254F8 (SPLNTUT02) | 529 | 1173 |
| | | | | GBI.g8576128.edit | 1 | 769 |
| | | | | 72038106V1 | 951 | 1892 |
| 41 | 5776350CB1 | 3172 | 1036-1253, 747-802, 82-257, 2389-3172, 1401-1437 | 71397725V1 | 1638 | 2301 |
| | | | | 7741938H1 (THYMNOE01) | 496 | 913 |
| | | | | GBI.g4034471.edit.1 () | 1 | 638 |
| | | | | 7741938J1 (THYMNOE01) | 798 | 1533 |
| | | | | 3400685H1 (UTRSNOT16) | 2579 | 2813 |
| | | | | g5836340 | 289 | 738 |
| | | | | 71164543V1 | 1693 | 2371 |
| | | | | 3992505T6 (LUNGNON03) | 2038 | 2650 |
| | | | | 71761861V1 | 955 | 1705 |
| | | | | 3042523F6 (HEAANOT01) | 2585 | 3172 |

Table 4 (cont.)

| Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID | Sequence Length | Selected Fragment(s) | Sequence Fragments | 5' Position | 3' Position |
|---|---|---|---|---|---|---|
| 42 | 7473300CB1 | 1997 | 1-467, 523-1997 | FL7473300CB1_00002 | 1 | 1997 |

Table 5

| Polynucleotide SEQ ID NO: | Incyte Project ID | Representative Library |
|---|---|---|
| 22 | 5155802CB1 | BONRFEC01 |
| 23 | 71269782CB1 | OVARTUT03 |
| 26 | 2759385CB1 | TESTTUT02 |
| 27 | 4226182CB1 | BRSTTUT02 |
| 28 | 5078962CB1 | BRABDIK02 |
| 29 | 7474340CB1 | TONSDIT01 |
| 30 | 7477287CB1 | PROSNOT06 |
| 31 | 2994162CB1 | HEAADIR01 |
| 32 | 3965293CB1 | PROSNOT14 |
| 33 | 4948403CB1 | PROSTMC01 |
| 34 | 7473165CB1 | BRAENOT02 |
| 35 | 7476667CB1 | EOSIHET02 |
| 37 | 3671788CB1 | PGANNOT01 |
| 38 | 7479181CB1 | PLACNOT02 |
| 39 | 6621372CB1 | THYRDIE01 |
| 40 | 4847254CB1 | SPLNTUT02 |
| 41 | 5776350CB1 | LUNGNON03 |

# Table 6

| Library | Vector | Library Description |
|---|---|---|
| BONRFEC01 | pINCY | This large size-fractionated library was constructed using RNA isolated from rib bone tissue removed from a Caucasian male fetus who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation. Serologies were negative. |
| BRABDIK02 | PSPORT1 | This amplified and normalized library was constructed using pooled cDNA from three different donors. cDNA was generated using mRNA isolated from diseased vermis tissue removed from a 79-year-old Caucasian female (donor A) who died from pneumonia, an 83-year-old Caucasian male (donor B) who died from congestive heart failure, and an 87-year-old Caucasian female (donor C) who died from esophageal cancer. Pathology indicated severe Alzheimer's disease in donors A & B and moderate Alzheimer's disease in donor C. Patient history included glaucoma, pseudophakia, gastritis with gastrointestinal bleeding, peripheral vascular disease, chronic obstructive pulmonary disease, seizures, tobacco abuse in remission, and transitory ischemic attacks in donor A; Parkinson's disease and atherosclerosis in donor B; hypertension, coronary artery disease, cerebral vascular accident, and hypothyroidism in donor C. Family history included Alzheimer's disease in the mother and sibling(s) of donor A. Independent clones from this amplified library were normalized in one round using conditions adapted Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used. |
| BRAENOT02 | pINCY | Library was constructed using RNA isolated from posterior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. |
| BRSTTUT02 | PSPORT1 | Library was constructed using RNA isolated from breast tumor tissue removed from a 54-year-old Caucasian female during a bilateral radical mastectomy with reconstruction. Pathology indicated residual invasive grade 3 mammary ductal adenocarcinoma. The remaining breast parenchyma exhibited proliferative fibrocystic changes without atypia. One of 10 axillary lymph nodes had metastatic tumor as a microscopic intranodal focus. Patient history included kidney infection and condyloma acuminatum. Family history included benign hypertension, hyperlipidemia, and a malignant colon neoplasm. |
| EOSIHET02 | PBLUESCRIPT | Library was constructed using RNA isolated from peripheral blood cells apheresed from a 48-year-old Caucasian male. Patient history included hypereosinophilia. The cell population was determined to be greater than 77% eosinophils by Wright's staining. |
| HEAADIR01 | pINCY | The library was constructed using RNA isolated from diseased right atrium and heart muscle wall tissue removed from a 7-month-old Caucasian male who died from cardiopulmonary arrest due to Pompe's disease. Patient history included Pompe's disease, left ventricular hypertrophy, pyrexia, right completec left lip, cleft palate, chronic serous otitis media, hypertrophic cardiomyopathy, congestive heart failure, and developmental delays. Family history included acute myocardial infarction, diabetes, cystic fibrosis, and Down's syndrome. |

Table 6 (cont.)

| Library | Vector | Library Description |
|---|---|---|
| LUNGNON03 | PSPORT1 | This normalized library was constructed from 2.56 million independent clones from a lung tissue library. RNA was made from lung tissue removed from the left lobe a 58-year-old Caucasian male during a segmental lung resection. Pathology for the associated tumor tissue indicated a metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Patient also received radiation therapy to the retroperitoneum. Family history included prostate cancer, breast cancer, and acute leukemia. The normalization and hybridization conditions were adapted from Soares et al., PNAS (1994) 91:9228; Swaroop et al., NAR (1991) 19:1954; and Bonaldo et al., Genome Research (1996) 6:791. |
| OVARTUT03 | pINCY | Library was constructed using RNA isolated from ovarian tumor tissue removed from the left ovary of a 52-year-old mixed ethnicity female during a total abdominal hysterectomy, bilateral salpingo-oophorectomy, peritoneal and lymphatic structure biopsy, regional lymph node excision, and peritoneal tissue destruction. Pathology indicated an invasive grade 3 (of 4) seroanaplastic carcinoma forming a mass in the left ovary. Multiple tumor implants were present on the surface of the left ovary and fallopian tube, right ovary and fallopian tube, posterior surface of the uterus, and cul-de-sac. The endometrium was atrophic. Multiple (2) leiomyomata were identified, one subserosal and 1 intramural. Pathology also indicated a metastatic grade 3 seroanaplastic carcinoma involving the omentum, cul-de-sac peritoneum, left broad ligament peritoneum, and mesentery colon. Patient history included breast cancer, chronic peptic ulcer, and joint pain. Family history included colon cancer, cerebrovascular disease, breast cancer, type II diabetes, esophagus cancer, and depressive disorder. |
| PGANNOT01 | PSPORT1 | Library was constructed using RNA isolated from paraganglionic tumor tissue removed from the intra-abdominal region of a 46-year-old Caucasian male during exploratory laparotomy. Pathology indicated a benign paraganglioma and was associated with a grade 2 renal cell carcinoma, clear cell type, which did not penetrate the capsule. Surgical margins were negative for tumor. |
| PLACNOT02 | pINCY | Library was constructed using RNA isolated from the placental tissue of a Hispanic female fetus, who was prematurely delivered at 21 weeks' gestation. Serologies of the mother's blood were positive for CMV (cytomegalovirus). |

# Table 6 (cont.)

| Library | Vector | Library Description |
|---|---|---|
| PROSNOT14 | pINCY | Library was constructed using RNA isolated from diseased prostate tissue removed from a 60-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). Patient history included a kidney cyst and hematuria. Family history included benign hypertension, cerebrovascular disease, and arteriosclerotic coronary artery disease. |
| PROSNOT06 | PSPORT | Library was constructed using RNA isolated from the diseased prostate tissue of a 57-year-old Caucasian male during radical prostatectomy, removal of both testes and excision of regional lymph nodes. Pathology indicated adenofibromatous hyperplasia. Pathology for the matched tumor tissue indicated adenocarcinoma (Gleason grade 3+3) in both the left and right periphery of the prostate. There was perineural invasion, and the tumor perforated the capsule. A single right pelvic lymph node and the right and left apical surgical margins were positive for tumor. Patient history included a benign neoplasm of the large bowel and type I diabetes. Patient medications included insulin. Family history included a malignant neoplasm of the prostate in the father and type I diabetes in the mother. |
| PROSTMC01 | pINCY | This size-selected library was constructed using RNA isolated from diseased prostate tissue removed from a 55-year-old Caucasian male during a radical prostatectomy, regional lymph node excision, and prostate needle biopsy. Pathology indicated adenofibromatous hyperplasia. Pathology for the matched tumor tissue indicated adenocarcinoma, Gleason grade 5+4, forming a predominant mass involving the left side peripherally with extension into the right posterior superior region. The tumor invaded and perforated the capsule to involve periprostatic tissue in the left posterior superior region. The left inferior and superior posterior surgical margins were positive. The right and left seminal vesicles, bladder neck tissue (after re-excision), and multiple pelvic lymph nodes were negative for tumor. One (of 9) left pelvic lymph nodes was metastatically involved. The patient presented with elevated prostate specific antigen (PSA). Patient history included calculus of the kidney. Previous surgeries included an adenotonsillectomy. Patient medications included Khats claw, an herbal preparation. Family history included breast cancer in the mother; lung cancer in the father; and breast cancer in the sibling(s). |
| SPLNTUT02 | pINCY | Library was constructed using RNA isolated from spleen tumor tissue obtained from a 45-year-old male during a staging laparotomy. Pathology indicated nodular sclerosing type of Hodgkin's disease forming innumerable nodules. Multiple lymph nodes were positive for Hodgkin's disease. |

Table 6 (cont.)

| Library | Vector | Library Description |
|---|---|---|
| TESTTUT02 | pINCY | Library was constructed using RNA isolated from testicular tumor removed from a 31-year-old Caucasian male during unilateral orchiectomy. Pathology indicated embryonal carcinoma. |
| THYRDIE01 | PCDNA2.1 | This 5' biased random primed library was constructed using RNA isolated from diseased thyroid tissue removed from a 22-year-old Caucasian female during closed thyroid biopsy, partial thyroidectomy, and regional lymph node excision. Pathology indicated adenomatous hyperplasia. The patient presented with malignant neoplasm of the thyroid. Patient history included normal delivery, alcohol abuse, and tobacco abuse. Previous surgeries included myringotomy. Patient medications included an unspecified type of birth control pills. Family history included hyperlipidemia and depressive disorder in the mother; and benign hypertension, congestive heart failure, and chronic leukemia in the grandparent(s). |
| TONSDIT01 | pINCY | Library was constructed using RNA isolated from the tonsil tissue of a 6-year-old Caucasian male during adenotonsillectomy. Pathology indicated lymphoid hyperplasia of the tonsils. The patient presented with an abscess of the pharynx. The patient was not taking any medications. Family history included hypothyroidism in the grandparent(s) and benign skin neoplasm in the sibling(s). |

# Table 7

| Program | Description | Reference | Parameter Threshold |
|---|---|---|---|
| ABI FACTURA | A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences. | Applied Biosystems, Foster City, CA. | |
| ABI/PARACEL FDF | A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences. | Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA. | Mismatch <50% |
| ABI AutoAssembler | A program that assembles nucleic acid sequences. | Applied Biosystems, Foster City, CA. | |
| BLAST | A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx. | Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402. | *ESTs:* Probability value= 1.0E-8 or less<br>*Full Length sequences*: Probability value= 1.0E-10 or less |
| FASTA | A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch. | Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489. | *ESTs:* fasta E value=1.06E-6<br>*Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less<br>*Full Length sequences:* fastx score=100 or greater |
| BLIMPS | A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions. | Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424. | Probability value= 1.0E-3 or less |
| HMMER | An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM. | Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350. | *PFAM hits:* Probability value= 1.0E-3 or less<br>*Signal peptide hits:* Score= 0 or greater |

# Table 7 (cont.)

| Program | Description | Reference | Parameter Threshold |
|---|---|---|---|
| ProfileScan | An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite. | Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221. | Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1. |
| Phred | A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability. | Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194. | |
| Phrap | A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences. | Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA. | Score= 120 or greater; Match length= 56 or greater |
| Consed | A graphical tool for viewing and editing Phrap assemblies. | Gordon, D. et al. (1998) Genome Res. 8:195-202. | |
| SPScan | A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides. | Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439. | Score=3.5 or greater |
| TMAP | A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation. | Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371. | |
| TMHMMER | A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation. | Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182. | |
| Motifs | A program that searches amino acid sequences for patterns that matched those defined in Prosite. | Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI. | |

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-21.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:22-42.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

18. A method for treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological

sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of PRTS in a biological sample comprising the steps of:

a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of PRTS in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of PRTS in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibodies from said animal; and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:22.

67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:23.

68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:24.

69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:25.

70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:26.

71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:29.

74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:30.

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:31.

76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:32.

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:33.

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:34.

79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:35.

80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:36.

81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:37.

82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:38.

83. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:39.

84. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:40.

85. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:41.

86. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:42.

<110> INCYTE GENOMICS, INC.
DELEGEANE, Angelo M.
GANDHI, Ameena R.
HAFALIA, April J.A.
LU, Dyung Aina M.
PATTERSON, Chandra
TRIBOULEY, Catherine M.
DAS, Debopriya
KALLICK Deborah A.
NGUYEN Danniel B.
Lee, Ernestine A.
KHAN Farrah A.
Yue, Henry
AU-YOUNG, Janice
GRIFFIN, Jennifer A.
POLICKY, Jennifer L.
RAMKUMAR, Jayalaxmi
YANG, Junming
THANGAVELU, Kavitha
DING, Li
KEARNEY, Liam
BAUGHN Mariah R.
BORROWSKY, Mark L.
SANJANWALA, Madhu S.
YAO, Monique G.
BURFORD, Neil
WALIA, Narinder K.
LAL, Preeti
LEE, Sally
TODD, Stephen
LO, Terrence P.
TANG, Y. Tom
ELLIOTT, Vicki S.
AZIMZAI, Yalda
LU, Yan

<120> PROTEASES

<130> PI-0167 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/220,063; 60/221,680; 60/223,544; 60/224,717; 60/225,988; 60/227,568
<151> 2000-07-21; 2000-07-28; 2000-08-04; 2000-08-11; 2000-08-16; 2000-08-23

<160> 42
<170> PERL Program

<210> 1
<211> 767
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 5155802CD1

<400> 1

```
Met Pro Thr Val Ile Ser Ala Ser Val Ala Pro Arg Thr Ala Ala
  1               5                  10                  15
Glu Pro Arg Ser Pro Gly Pro Val Pro His Pro Ala Gln Ser Lys
                20                  25                  30
Ala Thr Glu Ala Gly Gly Gly Asn Pro Ser Gly Ile Tyr Ser Ala
                35                  40                  45
Ile Ile Ser Arg Asn Phe Pro Ile Ile Gly Val Lys Glu Lys Thr
                50                  55                  60
Phe Glu Gln Leu His Lys Lys Cys Leu Glu Lys Lys Val Leu Tyr
                65                  70                  75
Val Asp Pro Glu Phe Pro Pro Asp Glu Thr Ser Leu Phe Tyr Ser
                80                  85                  90
Gln Lys Phe Pro Ile Gln Phe Val Trp Lys Arg Pro Pro Glu Ile
                95                 100                 105
Cys Glu Asn Pro Arg Phe Ile Ile Asp Gly Ala Asn Arg Thr Asp
               110                 115                 120
Ile Cys Gln Gly Glu Leu Gly Asp Cys Trp Phe Leu Ala Ala Ile
               125                 130                 135
Ala Cys Leu Thr Leu Asn Gln His Leu Leu Phe Arg Val Ile Pro
               140                 145                 150
His Asp Gln Ser Phe Ile Glu Asn Tyr Ala Gly Ile Phe His Phe
               155                 160                 165
Gln Phe Trp Arg Tyr Gly Glu Trp Val Asp Val Val Ile Asp Asp
               170                 175                 180
Cys Leu Pro Thr Tyr Asn Asn Gln Leu Val Phe Thr Lys Ser Asn
               185                 190                 195
His Arg Asn Glu Phe Trp Ser Ala Leu Leu Glu Lys Ala Tyr Ala
               200                 205                 210
Lys Leu His Gly Ser Tyr Glu Ala Leu Lys Gly Gly Asn Thr Thr
               215                 220                 225
Glu Ala Met Glu Asp Phe Thr Gly Gly Val Thr Glu Phe Phe Glu
               230                 235                 240
Ile Arg Asp Ala Pro Ser Asp Met Tyr Lys Ile Met Lys Lys Ala
               245                 250                 255
Ile Glu Arg Gly Ser Leu Met Gly Cys Ser Ile Asp Thr Ile Ile
               260                 265                 270
Pro Val Gln Tyr Glu Thr Arg Met Ala Cys Gly Leu Val Arg Gly
               275                 280                 285
His Ala Tyr Ser Val Thr Gly Leu Asp Glu Val Pro Phe Lys Gly
               290                 295                 300
Glu Lys Val Lys Leu Val Arg Leu Arg Asn Pro Trp Gly Gln Val
               305                 310                 315
Glu Trp Asn Gly Ser Trp Ser Asp Arg Trp Lys Asp Trp Ser Phe
               320                 325                 330
Val Asp Lys Asp Glu Lys Ala Arg Leu Gln His Gln Val Thr Glu
               335                 340                 345
Asp Gly Glu Phe Trp Met Ser Tyr Glu Asp Phe Ile Tyr His Phe
               350                 355                 360
Thr Lys Leu Glu Ile Cys Asn Leu Thr Ala Asp Ala Leu Gln Ser
               365                 370                 375
Asp Lys Leu Gln Thr Trp Thr Val Ser Val Asn Glu Gly Arg Trp
               380                 385                 390
Val Arg Gly Cys Ser Ala Gly Gly Cys Arg Asn Phe Pro Asp Thr
```

```
                395                 400                 405
Phe Trp Thr Asn Pro Gln Tyr Arg Leu Lys Leu Leu Glu Glu Asp
                410                 415                 420
Asp Asp Pro Asp Asp Ser Glu Val Ile Cys Ser Phe Leu Val Ala
                425                 430                 435
Leu Met Gln Lys Asn Arg Arg Lys Asp Arg Lys Leu Gly Ala Ser
                440                 445                 450
Leu Phe Thr Ile Gly Phe Ala Ile Tyr Glu Val Pro Lys Glu Met
                455                 460                 465
His Gly Asn Lys Gln His Leu Gln Lys Asp Phe Phe Leu Tyr Asn
                470                 475                 480
Ala Ser Lys Ala Arg Ser Lys Thr Tyr Ile Asn Met Arg Glu Val
                485                 490                 495
Ser Gln Arg Phe Arg Leu Pro Pro Ser Glu Tyr Val Ile Val Pro
                500                 505                 510
Ser Thr Tyr Glu Pro His Gln Glu Gly Glu Phe Ile Leu Arg Val
                515                 520                 525
Phe Ser Glu Lys Arg Asn Leu Ser Glu Glu Val Glu Asn Thr Ile
                530                 535                 540
Ser Val Asp Arg Pro Val Pro Ile Ile Phe Val Ser Asp Arg Ala
                545                 550                 555
Asn Ser Asn Lys Glu Leu Gly Val Asp Gln Glu Ser Glu Glu Gly
                560                 565                 570
Lys Gly Lys Thr Ser Pro Asp Lys Gln Lys Gln Ser Pro Gln Pro
                575                 580                 585
Gln Pro Gly Ser Ser Asp Gln Glu Ser Glu Glu Gln Gln Gln Phe
                590                 595                 600
Arg Asn Ile Phe Lys Gln Ile Ala Gly Asp Asp Met Glu Ile Cys
                605                 610                 615
Ala Asp Glu Leu Lys Lys Val Leu Asn Thr Val Val Asn Lys His
                620                 625                 630
Lys Asp Leu Lys Thr His Gly Phe Thr Leu Glu Ser Cys Arg Ser
                635                 640                 645
Met Ile Ala Leu Met Asp Thr Asp Gly Ser Gly Lys Leu Asn Leu
                650                 655                 660
Gln Glu Phe His His Leu Trp Asn Lys Ile Lys Ala Trp Gln Lys
                665                 670                 675
Ile Phe Lys His Tyr Asp Thr Asp Gln Ser Gly Thr Ile Asn Ser
                680                 685                 690
Tyr Glu Met Arg Asn Ala Val Asn Asp Ala Gly Phe His Leu Asn
                695                 700                 705
Asn Gln Leu Tyr Asp Ile Ile Thr Met Arg Tyr Ala Asp Lys His
                710                 715                 720
Met Asn Ile Asp Phe Asp Ser Phe Ile Cys Cys Phe Val Arg Leu
                725                 730                 735
Glu Gly Met Phe Arg Ala Phe His Ala Phe Asp Lys Asp Gly Asp
                740                 745                 750
Gly Ile Ile Lys Leu Asn Val Leu Glu Trp Leu Gln Leu Thr Met
                755                 760                 765
Tyr Ala
```

<210> 2
<211> 574
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71269782CD1

<400> 2

```
Met Gly Glu Asn Glu Ala Ser Leu Pro Asn Thr Ser Leu Gln Gly
  1               5                  10                  15
Lys Lys Met Ala Tyr Gln Lys Val His Ala Asp Gln Arg Ala Pro
             20                  25                  30
Gly His Ser Gln Tyr Leu Asp Asn Asp Asp Leu Gln Ala Thr Ala
             35                  40                  45
Leu Asp Leu Glu Trp Asp Met Glu Lys Glu Leu Glu Glu Ser Gly
             50                  55                  60
Phe Asp Gln Phe Gln Leu Asp Gly Ala Glu Asn Gln Asn Leu Gly
             65                  70                  75
His Ser Glu Thr Ile Asp Leu Asn Leu Asp Ser Ile Gln Pro Ala
             80                  85                  90
Thr Ser Pro Lys Gly Arg Phe Gln Arg Leu Gln Glu Glu Ser Asp
             95                 100                 105
Tyr Ile Thr His Tyr Thr Arg Ser Ala Pro Lys Ser Asn Arg Cys
            110                 115                 120
Asn Phe Cys His Val Leu Lys Ile Leu Cys Thr Ala Thr Ile Leu
            125                 130                 135
Phe Ile Phe Gly Ile Leu Ile Gly Tyr Tyr Val His Thr Asn Cys
            140                 145                 150
Pro Ser Asp Ala Pro Ser Ser Gly Thr Val Asp Pro Gln Leu Tyr
            155                 160                 165
Gln Glu Ile Leu Lys Thr Ile Gln Ala Glu Asp Ile Lys Lys Ser
            170                 175                 180
Phe Arg Asn Leu Val Gln Leu Tyr Lys Asn Glu Asp Asp Met Glu
            185                 190                 195
Ile Ser Lys Lys Ile Lys Thr Gln Trp Thr Ser Leu Gly Leu Glu
            200                 205                 210
Asp Val Gln Phe Val Asn Tyr Ser Val Leu Leu Asp Leu Pro Gly
            215                 220                 225
Pro Ser Pro Ser Thr Val Thr Leu Ser Ser Ser Gly Gln Cys Phe
            230                 235                 240
His Pro Asn Gly Gln Pro Cys Ser Glu Glu Ala Arg Lys Asp Ser
            245                 250                 255
Ser Gln Asp Leu Leu Tyr Ser Tyr Ala Ala Tyr Ser Ala Lys Gly
            260                 265                 270
Thr Leu Lys Ala Glu Val Ile Asp Val Ser Tyr Gly Met Ala Asp
            275                 280                 285
Asp Leu Lys Arg Ile Arg Lys Ile Lys Asn Val Thr Asn Gln Ile
            290                 295                 300
Ala Leu Leu Lys Leu Gly Lys Leu Pro Leu Leu Tyr Lys Leu Ser
            305                 310                 315
Ser Leu Glu Lys Ala Gly Phe Gly Gly Val Leu Leu Tyr Ile Asp
            320                 325                 330
Pro Cys Asp Leu Pro Lys Thr Val Asn Pro Ser His Asp Thr Phe
            335                 340                 345
Met Val Ser Leu Asn Pro Gly Gly Asp Pro Ser Thr Pro Gly Tyr
            350                 355                 360
Pro Ser Val Asp Glu Ser Phe Arg Gln Ser Arg Ser Asn Leu Thr
            365                 370                 375
Ser Leu Leu Val Gln Pro Ile Ser Ala Ser Leu Val Ala Lys Leu
```

```
                380                 385                 390
Ile Ser Ser Pro Lys Ala Arg Thr Lys Asn Glu Ala Cys Ser Ser
                395                 400                 405
Leu Glu Leu Pro Asn Asn Glu Ile Arg Val Val Ser Met Gln Val
                410                 415                 420
Gln Thr Val Thr Lys Leu Lys Thr Val Thr Asn Val Val Gly Phe
                425                 430                 435
Val Met Gly Leu Thr Ser Pro Asp Arg Tyr Ile Ile Val Gly Ser
                440                 445                 450
His His His Thr Ala His Ser Tyr Asn Gly Gln Glu Trp Ala Ser
                455                 460                 465
Ser Thr Ala Ile Ile Thr Ala Phe Ile Arg Ala Leu Met Ser Lys
                470                 475                 480
Val Lys Arg Gly Trp Arg Pro Asp Arg Thr Ile Val Phe Cys Ser
                485                 490                 495
Trp Gly Gly Thr Ala Phe Gly Asn Ile Gly Ser Tyr Glu Trp Gly
                500                 505                 510
Glu Asp Phe Lys Lys Val Leu Gln Lys Asn Val Val Ala Tyr Ile
                515                 520                 525
Ser Leu His Ser Pro Ile Arg Gly Asn Ser Ser Leu Tyr Pro Val
                530                 535                 540
Ala Ser Pro Ser Leu Gln Gln Leu Val Val Glu Val Arg Gln Thr
                545                 550                 555
Thr Ile Val Ser Asn Asp Tyr Ala Lys Pro Thr Phe Ser Leu Tyr
                560                 565                 570
Phe Asp Ile Ser
```

<210> 3
<211> 320
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472651CD1

<400> 3

```
Met Gly Asp Pro Glu Gly Ser Ala Glu Trp Gly Trp Gly Lys Gly
  1                 5                  10                  15
Ile Pro Val Val Arg Arg Asn Leu Leu Thr Val Asp Gly Ile Ser
                 20                  25                  30
Leu Cys Leu Glu Gly Ser Trp Trp Arg Gln Lys Gly Pro Ala Ser
                 35                  40                  45
Pro Gly Phe Ser His Ser Leu Pro Arg Leu Gln Pro Asn Pro Gly
                 50                  55                  60
Pro Ser Ser Thr Met Trp Leu Leu Leu Thr Leu Ser Phe Leu Leu
                 65                  70                  75
Ala Ser Thr Ala Ala Gln Asp Gly Asp Lys Leu Leu Glu Gly Asp
                 80                  85                  90
Glu Cys Ala Pro His Ser Gln Pro Trp Gln Val Ala Leu Tyr Glu
                 95                 100                 105
Arg Gly Arg Phe Asn Cys Gly Ala Ser Leu Ile Ser Pro His Trp
                110                 115                 120
Val Leu Ser Ala Ala His Cys Gln Ser Arg Phe Met Arg Val Arg
                125                 130                 135
```

```
Leu Gly Glu His Asn Leu Arg Lys Arg Asp Gly Pro Glu Gln Leu
                140                 145                 150
Arg Thr Thr Ser Arg Val Ile Pro His Pro Arg Tyr Glu Ala Arg
                155                 160                 165
Ser His Arg Asn Asp Ile Met Leu Leu Arg Leu Val Gln Pro Ala
                170                 175                 180
Arg Leu Asn Pro Gln Val Arg Pro Ala Val Leu Pro Thr Arg Cys
                185                 190                 195
Pro His Pro Gly Glu Ala Cys Val Val Ser Gly Trp Gly Leu Val
                200                 205                 210
Ser His Asn Glu Pro Gly Thr Ala Gly Ser Pro Arg Ser Gln Val
                215                 220                 225
Ser Leu Pro Asp Thr Leu His Cys Ala Asn Ile Ser Ile Ile Ser
                230                 235                 240
Asp Thr Ser Cys Asp Lys Ser Tyr Pro Gly Arg Leu Thr Asn Thr
                245                 250                 255
Met Val Cys Ala Gly Ala Glu Gly Arg Gly Ala Glu Ser Cys Glu
                260                 265                 270
Gly Asp Ser Gly Gly Pro Leu Val Cys Gly Gly Ile Leu Gln Gly
                275                 280                 285
Ile Val Ser Trp Gly Asp Val Pro Cys Asp Asn Thr Thr Lys Pro
                290                 295                 300
Gly Val Tyr Thr Lys Val Cys His Tyr Leu Glu Trp Ile Arg Glu
                305                 310                 315
Thr Met Lys Arg Asn
                320
```

<210> 4
<211> 378
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7478251CD1

<400> 4

```
Met Ala Glu Lys Pro Ser Asn Gly Val Leu Val His Met Val Lys
  1                 5                  10                  15
Leu Leu Ile Lys Thr Phe Leu Asp Gly Ile Phe Asp Asp Leu Met
                 20                  25                  30
Glu Asn Asn Val Leu Asn Thr Asp Glu Ile His Leu Ile Gly Lys
                 35                  40                  45
Cys Leu Lys Phe Val Val Ser Asn Ala Glu Asn Leu Val Asp Asp
                 50                  55                  60
Ile Thr Glu Thr Ala Gln Thr Ala Gly Lys Ile Phe Arg Glu His
                 65                  70                  75
Leu Trp Asn Ser Lys Lys Gln Leu Ser Ser Ile Phe Phe Ser Leu
                 80                  85                  90
Ser Ala Phe Leu Glu Ile Gln Gly Ala Gln Pro Ser Gly Lys Leu
                 95                 100                 105
Lys Leu Cys Pro His Ala His Phe His Glu Leu Lys Thr Lys Arg
                110                 115                 120
Ala Asp Glu Ile Tyr Pro Val Met Glu Lys Lys Arg Arg Thr Cys
                125                 130                 135
Leu Gly Leu Asn Ile Arg Asn Lys Glu Phe Asn Tyr Leu His Asn
```

```
                140                 145                 150
Arg Asn Gly Ser Glu Leu Asp Leu Leu Gly Met Arg Asp Leu Leu
                155                 160                 165
Glu Asn Leu Gly Tyr Ser Val Val Ile Lys Glu Asn Leu Thr Ala
                170                 175                 180
Gln Glu Met Glu Thr Ala Leu Arg Gln Phe Ala Ala His Pro Glu
                185                 190                 195
His Gln Ser Ser Asp Ser Thr Phe Leu Val Phe Met Ser His Ser
                200                 205                 210
Ile Leu Asn Gly Ile Cys Gly Thr Lys His Trp Asp Gln Glu Pro
                215                 220                 225
Asp Val Leu His Asp Asp Thr Ile Phe Glu Ile Phe Asn Asn Arg
                230                 235                 240
Asn Cys Gln Ser Leu Lys Asp Lys Pro Lys Val Ile Ile Met Gln
                245                 250                 255
Ala Cys Arg Gly Asn Gly Ala Gly Ile Val Trp Phe Thr Thr Asp
                260                 265                 270
Ser Gly Lys Ala Gly Ala Asp Thr His Gly Arg Leu Leu Gln Gly
                275                 280                 285
Asn Ile Cys Asn Asp Ala Val Thr Lys Ala His Val Glu Lys Asp
                290                 295                 300
Phe Ile Ala Phe Lys Ser Ser Thr Pro His Asn Val Ser Trp Arg
                305                 310                 315
His Glu Thr Asn Gly Ser Val Phe Ile Ser Gln Ile Ile Tyr Tyr
                320                 325                 330
Phe Arg Glu Tyr Ser Trp Ser His His Leu Glu Glu Ile Phe Gln
                335                 340                 345
Lys Val Gln His Ser Phe Glu Thr Pro Asn Ile Leu Thr Gln Leu
                350                 355                 360
Pro Thr Ile Glu Arg Leu Ser Met Thr Arg Tyr Phe Tyr Leu Phe
                365                 370                 375
Pro Gly Asn
```

<210> 5
<211> 366
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2759385CD1

<400> 5

```
Met Thr Val Arg Asn Ile Ala Ser Ile Cys Asn Met Gly Thr Asn
  1                 5                  10                  15
Ala Ser Ala Leu Glu Lys Asp Ile Gly Pro Glu Gln Phe Pro Ile
                 20                  25                  30
Asn Glu His Tyr Phe Gly Leu Val Asn Phe Gly Asn Thr Cys Tyr
                 35                  40                  45
Cys Asn Ser Val Leu Gln Ala Leu Tyr Phe Cys Arg Pro Phe Arg
                 50                  55                  60
Glu Asn Val Leu Ala Tyr Lys Ala Gln Gln Lys Lys Lys Glu Asn
                 65                  70                  75
Leu Leu Thr Cys Leu Ala Asp Leu Phe His Ser Ile Ala Thr Gln
                 80                  85                  90
```

```
Lys Lys Lys Val Gly Val Ile Pro Pro Lys Lys Phe Ile Ser Arg
                95                  100                 105
Leu Arg Lys Glu Asn Asp Leu Phe Asp Asn Tyr Met Gln Gln Asp
            110                 115                 120
Ala His Glu Phe Leu Asn Tyr Leu Leu Asn Thr Ile Ala Asp Ile
            125                 130                 135
Leu Gln Glu Glu Lys Lys Gln Glu Lys Gln Asn Gly Lys Leu Lys
            140                 145                 150
Asn Gly Asn Met Asn Glu Pro Ala Glu Asn Asn Lys Pro Glu Leu
            155                 160                 165
Thr Trp Val His Glu Ile Phe Gln Gly Thr Leu Thr Asn Glu Thr
            170                 175                 180
Arg Cys Leu Asn Cys Glu Thr Val Ser Ser Lys Asp Glu Asp Phe
            185                 190                 195
Leu Asp Leu Ser Val Asp Val Glu Gln Asn Thr Ser Ile Thr His
            200                 205                 210
Cys Leu Arg Asp Phe Ser Asn Thr Glu Thr Leu Cys Ser Glu Gln
            215                 220                 225
Lys Tyr Tyr Cys Glu Thr Cys Cys Ser Lys Gln Glu Ala Gln Lys
            230                 235                 240
Arg Met Arg Val Lys Lys Leu Pro Met Ile Leu Ala Leu His Leu
            245                 250                 255
Lys Arg Phe Lys Tyr Met Glu Gln Leu His Arg Tyr Thr Lys Leu
            260                 265                 270
Ser Tyr Arg Val Val Phe Pro Leu Glu Leu Arg Leu Phe Asn Thr
            275                 280                 285
Ser Ser Asp Ala Val Asn Leu Asp Arg Met Tyr Asp Leu Val Ala
            290                 295                 300
Val Val Val His Cys Gly Ser Gly Pro Asn Arg Gly His Tyr Ile
            305                 310                 315
Thr Ile Val Lys Ser His Gly Phe Trp Leu Leu Phe Asp Asp Asp
            320                 325                 330
Ile Val Glu Lys Ile Asp Ala Gln Ala Ile Glu Glu Phe Tyr Gly
            335                 340                 345
Leu Thr Ser Asp Ile Ser Lys Asn Ser Glu Ser Gly Tyr Ile Leu
            350                 355                 360
Phe Tyr Gln Ser Arg Glu
            365

<210> 6
<211> 389
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4226182CD1

<400> 6
Met Asp Tyr Pro Arg Tyr Leu Gly Ala Val Phe Pro Gly Thr Met
  1                 5                  10                  15
Cys Ile Thr Arg Tyr Ser Ala Gly Val Ala Leu Gln Cys Gly Pro
            20                  25                  30
Ala Ser Cys Cys Asp Phe Arg Thr Cys Val Leu Lys Asp Gly Ala
            35                  40                  45
Lys Cys Tyr Lys Gly Leu Cys Cys Lys Asp Cys Gln Ile Leu Gln
```

```
                50                  55                  60
Ser Gly Val Glu Cys Arg Pro Lys Ala His Pro Glu Cys Asp Ile
                65                  70                  75
Ala Glu Asn Cys Asn Gly Ser Ser Pro Glu Cys Gly Pro Asp Ile
                80                  85                  90
Thr Leu Ile Asn Gly Leu Ser Cys Lys Asn Asn Lys Phe Ile Cys
                95                 100                 105
Tyr Asp Gly Asp Cys His Asp Leu Asp Ala Arg Cys Glu Ser Val
               110                 115                 120
Phe Gly Lys Gly Ser Arg Asn Ala Pro Phe Ala Cys Tyr Glu Glu
               125                 130                 135
Ile Gln Ser Gln Ser Asp Arg Phe Gly Asn Cys Gly Arg Asp Arg
               140                 145                 150
Asn Asn Lys Tyr Val Phe Cys Gly Trp Arg Asn Leu Ile Cys Gly
               155                 160                 165
Arg Leu Val Cys Thr Tyr Pro Thr Arg Lys Pro Phe His Gln Glu
               170                 175                 180
Asn Gly Asp Val Ile Tyr Ala Phe Val Arg Asp Ser Val Cys Ile
               185                 190                 195
Thr Val Asp Tyr Lys Leu Pro Arg Thr Val Pro Asp Pro Leu Ala
               200                 205                 210
Val Lys Asn Gly Ser Gln Cys Asp Ile Gly Arg Val Cys Val Asn
               215                 220                 225
Arg Glu Cys Val Glu Ser Arg Ile Ile Lys Ala Ser Ala His Val
               230                 235                 240
Cys Ser Gln Gln Cys Ser Gly His Gly Val Cys Asp Ser Arg Asn
               245                 250                 255
Lys Cys His Cys Ser Pro Gly Tyr Lys Pro Pro Asn Cys Gln Ile
               260                 265                 270
Arg Ser Lys Gly Phe Ser Ile Phe Pro Glu Glu Asp Met Gly Ser
               275                 280                 285
Ile Met Glu Arg Ala Ser Gly Lys Thr Glu Asn Thr Trp Leu Leu
               290                 295                 300
Gly Phe Leu Ile Ala Leu Pro Ile Leu Ile Val Thr Thr Ala Ile
               305                 310                 315
Val Leu Ala Arg Lys Gln Leu Lys Lys Trp Phe Ala Lys Glu Glu
               320                 325                 330
Glu Phe Pro Ser Ser Glu Ser Lys Ser Glu Gly Ser Thr Gln Thr
               335                 340                 345
Tyr Ala Ser Gln Ser Ser Ser Glu Gly Ser Thr Gln Thr Tyr Ala
               350                 355                 360
Ser Gln Thr Arg Ser Glu Ser Ser Ser Gln Ala Asp Thr Ser Lys
               365                 370                 375
Ser Lys Ser Gln Asp Ser Thr Gln Thr Gln Ser Ser Ser Asn
               380                 385

<210> 7
<211> 217
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5078962CD1

<400> 7
```

```
Met Thr Thr Glu Glu Ile Asp Ala Leu Val His Arg Glu Ile Ile
  1               5                  10                  15
Ser His Asn Ala Tyr Pro Ser Pro Leu Gly Tyr Gly Gly Phe Pro
                20                  25                  30
Lys Ser Val Cys Thr Ser Val Asn Asn Val Leu Cys His Gly Ile
                35                  40                  45
Pro Asp Ser Arg Pro Leu Gln Asp Gly Asp Ile Ile Asn Ile Asp
                50                  55                  60
Val Thr Val Tyr Tyr Asn Gly Tyr His Gly Asp Thr Ser Glu Thr
                65                  70                  75
Phe Leu Val Gly Asn Val Asp Glu Cys Gly Lys Lys Leu Val Glu
                80                  85                  90
Val Ala Arg Arg Cys Arg Asp Glu Ala Ile Ala Ala Cys Arg Ala
                95                 100                 105
Gly Ala Pro Phe Ser Val Ile Gly Asn Thr Ile Ser His Ile Thr
               110                 115                 120
His Gln Asn Gly Phe Gln Val Cys Pro His Phe Val Gly His Gly
               125                 130                 135
Ile Gly Ser Tyr Phe His Gly His Pro Glu Ile Trp His His Ala
               140                 145                 150
Asn Asp Ser Asp Leu Pro Met Glu Glu Gly Met Ala Phe Thr Ile
               155                 160                 165
Glu Pro Ile Ile Thr Glu Gly Ser Pro Glu Phe Lys Val Leu Glu
               170                 175                 180
Asp Ala Trp Thr Val Val Ser Leu Asp Asn Gln Arg Ser Ala Gln
               185                 190                 195
Phe Glu His Thr Val Leu Ile Thr Ser Arg Gly Ala Gln Ile Leu
               200                 205                 210
Thr Lys Leu Pro His Glu Ala
               215
```

<210> 8
<211> 486
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474340CD1

<400> 8

```
Met Glu Arg Asp Ser His Gly Asn Ala Ser Pro Ala Arg Thr Pro
  1               5                  10                  15
Ser Ala Gly Ala Ser Pro Ala Gln Ala Ser Pro Ala Gly Thr Pro
                20                  25                  30
Pro Gly Arg Ala Ser Pro Ala Gln Ala Ser Pro Ala Gln Ala Ser
                35                  40                  45
Pro Ala Gly Thr Pro Pro Gly Arg Ala Ser Pro Ala Gln Ala Ser
                50                  55                  60
Pro Ala Gly Thr Pro Pro Gly Arg Ala Ser Pro Gly Arg Ala Ser
                65                  70                  75
Pro Ala Gln Ala Ser Pro Ala Arg Ala Ser Pro Ala Leu Ala Ser
                80                  85                  90
Leu Ser Arg Ser Ser Ser Gly Arg Ser Ser Ser Ala Arg Ser Ala
                95                 100                 105
Ser Val Thr Thr Ser Pro Thr Arg Val Tyr Leu Val Arg Ala Thr
```

```
                110                 115                 120
Pro Val Gly Ala Val Pro Ile Arg Ser Ser Pro Ala Arg Ser Ala
                125                 130                 135
Pro Ala Thr Arg Ala Thr Arg Glu Ser Pro Gly Thr Ser Leu Pro
                140                 145                 150
Lys Phe Thr Trp Arg Glu Gly Gln Lys Gln Leu Pro Leu Ile Gly
                155                 160                 165
Cys Val Leu Leu Leu Ile Ala Leu Val Val Ser Leu Ile Ile Leu
                170                 175                 180
Phe Gln Phe Trp Gln Gly His Thr Gly Ile Arg Tyr Lys Glu Gln
                185                 190                 195
Arg Glu Ser Cys Pro Lys His Ala Val Arg Cys Asp Gly Val Val
                200                 205                 210
Asp Cys Lys Leu Lys Ser Asp Glu Leu Gly Cys Val Arg Phe Asp
                215                 220                 225
Trp Asp Lys Ser Leu Leu Lys Ile Tyr Ser Gly Ser Ser His Gln
                230                 235                 240
Trp Leu Pro Ile Cys Ser Ser Asn Trp Asn Asp Ser Tyr Ser Glu
                245                 250                 255
Lys Thr Cys Gln Gln Leu Gly Phe Glu Ser Ala His Arg Thr Thr
                260                 265                 270
Glu Val Ala His Arg Asp Phe Ala Asn Ser Phe Ser Ile Leu Arg
                275                 280                 285
Tyr Asn Ser Thr Ile Gln Glu Ser Leu His Arg Ser Glu Cys Pro
                290                 295                 300
Ser Gln Arg Tyr Ile Ser Leu Gln Cys Ser His Cys Gly Leu Arg
                305                 310                 315
Ala Met Thr Gly Arg Ile Val Gly Gly Ala Leu Ala Ser Asp Ser
                320                 325                 330
Lys Trp Pro Trp Gln Val Ser Leu His Phe Gly Thr Thr His Ile
                335                 340                 345
Cys Gly Gly Thr Leu Ile Asp Ala Gln Trp Val Leu Thr Ala Ala
                350                 355                 360
His Cys Phe Phe Val Thr Arg Glu Lys Val Leu Glu Gly Trp Lys
                365                 370                 375
Val Tyr Ala Gly Thr Ser Asn Leu His Gln Leu Pro Glu Ala Ala
                380                 385                 390
Ser Ile Ala Glu Ile Ile Ile Asn Ser Asn Tyr Thr Asp Glu Glu
                395                 400                 405
Asp Asp Tyr Asp Ile Ala Leu Met Arg Leu Ser Lys Pro Leu Thr
                410                 415                 420
Leu Ser Gly Glu Gly Ile Cys Thr Pro Arg Ser Pro Ala Pro Gln
                425                 430                 435
Pro Gln His Pro Leu Gln Pro Ser His Leu Ser Ala Ser Val Asn
                440                 445                 450
Ser Tyr Pro Gly Pro Lys Ala Ser Ala Gly Gln Lys Ser Lys Thr
                455                 460                 465
Leu Lys Asp Pro Tyr Met Glu His Phe Cys Phe Ile Ile Arg Glu
                470                 475                 480
Thr Glu Ala Gln Gly Leu
                485
```

<210> 9
<211> 390
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477287CD1

<400> 9
Met Gly Pro Arg Leu Ile Pro Phe Leu Phe Leu Phe Val Tyr Pro
1 5 10 15
Ile Leu Cys Arg Ile Ile Leu Arg Lys Gly Lys Ser Ile Arg Gln
20 25 30
Arg Met Glu Glu Gln Gly Val Leu Glu Thr Phe Leu Arg Asp His
35 40 45
Pro Lys Ala Asp Pro Ile Ala Lys Tyr Tyr Phe Asn Asn Asp Ala
50 55 60
Val Ala Tyr Glu Pro Phe Thr Asn Tyr Leu Asp Ser Phe Tyr Phe
65 70 75
Gly Glu Ile Ser Thr Gly Thr Pro Pro Gln Asn Phe Leu Val Ser
80 85 90
Leu Ile Arg Val Pro Pro Ile Cys Ser Leu Pro Ser Ile Tyr Cys
95 100 105
Gln Ser Gln Val Cys Ser Asn His Asn Arg Phe Asn Pro Ser Leu
110 115 120
Ser Ser Thr Phe Arg Asn Asp Gly Gln Thr Tyr Gly Leu Ser Tyr
125 130 135
Gly Ser Gly Ser Leu Ser Val Phe Leu Gly Tyr Asp Thr Val Thr
140 145 150
Val His Asn Ile Val Val Asn Asn Gln Glu Phe Gly Leu Ser Glu
155 160 165
Asn Glu Pro Ser Asp Pro Phe Tyr Tyr Ser Asp Phe Asp Gly Ile
170 175 180
Leu Gly Met Ala Tyr Pro Asn Met Ala Glu Gly Asn Ser Pro Thr
185 190 195
Val Met Gln Gly Met Leu Gln Gln Ser Gln Leu Thr Gln Pro Val
200 205 210
Phe Ser Phe Tyr Phe Thr Cys Gln Pro Thr Arg Gln Tyr Cys Gly
215 220 225
Glu Leu Ile Leu Gly Gly Val Asp Pro Asn Leu Tyr Ser Gly Gln
230 235 240
Ile Ile Trp Thr Pro Val Ser Pro Glu Leu Tyr Trp Gln Ile Ala
245 250 255
Ile Glu Glu Phe Ala Ile Gly Asn Gln Ala Thr Gly Leu Cys Ser
260 265 270
Glu Gly Cys Gln Ala Ile Val Asp Thr Glu Thr Phe Leu Leu Ala
275 280 285
Val Pro Gln Gln Tyr Met Ala Ser Phe Leu Gln Ala Thr Gly Pro
290 295 300
Gln Gln Ala Gln Asn Gly Asp Phe Val Val Asn Cys Ser Tyr Ile
305 310 315
Gln Ser Met Pro Thr Ile Thr Phe Ile Ile Gly Gly Ala Gln Phe
320 325 330
Pro Leu Pro Pro Ser Glu Tyr Val Phe Asn Asn Asn Gly Tyr Cys
335 340 345
Arg Leu Gly Thr Glu Ala Thr Cys Leu Pro Ser Arg Ser Gly Gln
350 355 360
Pro Leu Trp Ile Leu Gly Asp Val Phe Leu Lys Glu Tyr Cys Ser
365 370 375
Val Tyr Asp Met Ala Asn Asn Arg Val Gly Phe Ala Phe Ser Ala

380 385 390

<210> 10
<211> 1916
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2994162CD1

<400> 10
Met Gly Ser Pro Asp Ala Ala Ala Ala Val Arg Lys Asp Arg Leu
1 5 10 15
His Pro Arg Gln Val Lys Leu Leu Glu Thr Leu Ser Glu Tyr Glu
20 25 30
Ile Val Ser Pro Ile Arg Val Asn Ala Leu Gly Glu Pro Phe Pro
35 40 45
Thr Asn Val His Phe Lys Arg Thr Arg Arg Ser Ile Asn Ser Ala
50 55 60
Thr Asp Pro Trp Pro Ala Phe Ala Ser Ser Ser Ser Ser Ser Thr
65 70 75
Ser Ser Gln Ala His Tyr Arg Leu Ser Ala Phe Gly Gln Gln Phe
80 85 90
Leu Phe Asn Leu Thr Ala Asn Ala Gly Phe Ile Ala Pro Leu Phe
95 100 105
Thr Val Thr Leu Leu Gly Thr Pro Gly Val Asn Gln Thr Lys Phe
110 115 120
Tyr Ser Glu Glu Glu Ala Glu Leu Lys His Cys Phe Tyr Lys Gly
125 130 135
Tyr Val Asn Thr Asn Ser Glu His Thr Ala Val Ile Ser Leu Cys
140 145 150
Ser Gly Met Leu Gly Thr Phe Arg Ser His Asp Gly Asp Tyr Phe
155 160 165
Ile Glu Pro Leu Gln Ser Met Asp Glu Gln Glu Asp Glu Glu Glu
170 175 180
Gln Asn Lys Pro His Ile Ile Tyr Arg Arg Ser Ala Pro Gln Arg
185 190 195
Glu Pro Ser Thr Gly Arg His Ala Cys Asp Thr Ser Glu His Lys
200 205 210
Asn Arg His Ser Lys Asp Lys Lys Lys Thr Arg Ala Arg Lys Trp
215 220 225
Gly Glu Arg Ile Asn Leu Ala Gly Asp Val Ala Ala Leu Asn Ser
230 235 240
Gly Leu Ala Thr Glu Ala Phe Ser Ala Tyr Gly Asn Lys Thr Asp
245 250 255
Asn Thr Arg Glu Lys Arg Thr His Arg Arg Thr Lys Arg Phe Leu
260 265 270
Ser Tyr Pro Arg Phe Val Glu Val Leu Val Val Ala Asp Asn Arg
275 280 285
Met Val Ser Tyr His Gly Glu Asn Leu Gln His Tyr Ile Leu Thr
290 295 300
Leu Met Ser Ile Val Ala Ser Ile Tyr Lys Asp Pro Ser Ile Gly
305 310 315
Asn Leu Ile Asn Ile Val Ile Val Asn Leu Ile Val Ile His Asn

```
                320                 325                 330
Glu Gln Asp Gly Pro Ser Ile Ser Phe Asn Ala Gln Thr Thr Leu
                335                 340                 345
Lys Asn Phe Cys Gln Trp Gln His Ser Lys Asn Ser Pro Gly Gly
                350                 355                 360
Ile His His Asp Thr Ala Val Leu Leu Thr Arg Gln Asp Ile Cys
                365                 370                 375
Arg Ala His Asp Lys Cys Asp Thr Leu Gly Leu Ala Glu Leu Gly
                380                 385                 390
Thr Ile Cys Asp Pro Tyr Arg Ser Cys Ser Ile Ser Glu Asp Ser
                395                 400                 405
Gly Leu Ser Thr Ala Phe Thr Ile Ala His Glu Leu Gly His Val
                410                 415                 420
Phe Asn Met Pro His Asp Asp Asn Asn Lys Cys Lys Glu Glu Gly
                425                 430                 435
Val Lys Ser Pro Gln His Val Met Ala Pro Thr Leu Asn Phe Tyr
                440                 445                 450
Thr Asn Pro Trp Met Trp Ser Lys Cys Ser Arg Lys Tyr Ile Thr
                455                 460                 465
Glu Phe Leu Asp Thr Gly Tyr Gly Glu Cys Leu Leu Asn Glu Pro
                470                 475                 480
Glu Ser Arg Pro Tyr Pro Leu Pro Val Gln Leu Pro Gly Ile Leu
                485                 490                 495
Tyr Asn Val Asn Lys Gln Cys Glu Leu Ile Phe Gly Pro Gly Ser
                500                 505                 510
Gln Val Cys Pro Tyr Met Met Gln Cys Arg Arg Leu Trp Cys Asn
                515                 520                 525
Asn Val Asn Gly Val His Lys Gly Cys Arg Thr Gln His Thr Pro
                530                 535                 540
Trp Ala Asp Gly Thr Glu Cys Glu Pro Gly Lys His Cys Lys Tyr
                545                 550                 555
Gly Phe Cys Val Pro Lys Glu Met Asp Val Pro Val Thr Asp Gly
                560                 565                 570
Ser Trp Gly Ser Trp Ser Pro Phe Gly Thr Cys Ser Arg Thr Cys
                575                 580                 585
Gly Gly Gly Ile Lys Thr Ala Ile Arg Glu Cys Asn Arg Pro Glu
                590                 595                 600
Pro Lys Asn Gly Gly Lys Tyr Cys Val Gly Arg Arg Met Lys Phe
                605                 610                 615
Lys Ser Cys Asn Thr Glu Pro Cys Leu Lys Gln Lys Arg Asp Phe
                620                 625                 630
Arg Asp Glu Gln Cys Ala His Phe Asp Gly Lys His Phe Asn Ile
                635                 640                 645
Asn Gly Leu Leu Pro Asn Val Arg Trp Val Pro Lys Tyr Ser Gly
                650                 655                 660
Ile Leu Met Lys Asp Arg Cys Lys Leu Phe Cys Arg Val Ala Gly
                665                 670                 675
Asn Thr Ala Tyr Tyr Gln Leu Arg Asp Arg Val Ile Asp Gly Thr
                680                 685                 690
Pro Cys Gly Gln Asp Thr Asn Asp Ile Cys Val Gln Gly Leu Cys
                695                 700                 705
Arg Gln Ala Gly Cys Asp His Val Leu Asn Ser Lys Ala Arg Arg
                710                 715                 720
Asp Lys Cys Gly Val Cys Gly Gly Asp Asn Ser Ser Cys Lys Thr
                725                 730                 735
Val Ala Gly Thr Phe Asn Thr Val His Tyr Gly Tyr Asn Thr Val
```

```
                740                 745                 750
Val Arg Ile Pro Ala Gly Ala Thr Asn Ile Asp Val Arg Gln His
                755                 760                 765
Ser Phe Ser Gly Glu Thr Asp Asp Asp Asn Tyr Leu Ala Leu Ser
                770                 775                 780
Ser Ser Lys Gly Glu Phe Leu Leu Asn Gly Asn Phe Val Val Thr
                785                 790                 795
Met Ala Lys Arg Glu Ile Arg Ile Gly Asn Ala Val Val Glu Tyr
                800                 805                 810
Ser Gly Ser Glu Thr Ala Val Glu Arg Ile Asn Ser Thr Asp Arg
                815                 820                 825
Ile Glu Gln Glu Leu Leu Leu Gln Val Leu Ser Val Gly Lys Leu
                830                 835                 840
Tyr Asn Pro Asp Val Arg Tyr Ser Phe Asn Ile Pro Ile Glu Asp
                845                 850                 855
Lys Pro Gln Gln Phe Tyr Trp Asn Ser His Gly Pro Trp Gln Ala
                860                 865                 870
Cys Ser Lys Pro Cys Gln Gly Glu Arg Lys Arg Lys Leu Val Cys
                875                 880                 885
Thr Arg Glu Ser Asp Gln Leu Thr Val Ser Asp Gln Arg Cys Asp
                890                 895                 900
Arg Leu Pro Gln Pro Gly His Ile Thr Glu Pro Cys Gly Thr Asp
                905                 910                 915
Cys Asp Leu Arg Trp His Val Ala Ser Arg Ser Glu Cys Ser Ala
                920                 925                 930
Gln Cys Gly Leu Gly Tyr Arg Thr Leu Asp Ile Tyr Cys Ala Lys
                935                 940                 945
Tyr Ser Arg Leu Asp Gly Lys Thr Glu Lys Val Asp Asp Gly Phe
                950                 955                 960
Cys Ser Ser His Pro Lys Pro Ser Asn Arg Glu Lys Cys Ser Gly
                965                 970                 975
Glu Cys Asn Thr Gly Gly Trp Arg Tyr Ser Ala Trp Thr Glu Cys
                980                 985                 990
Ser Lys Ser Cys Asp Gly Gly Thr Gln Arg Arg Arg Ala Ile Cys
                995                1000                1005
Val Asn Thr Arg Asn Asp Val Leu Asp Asp Ser Lys Cys Thr His
               1010                1015                1020
Gln Glu Lys Val Thr Ile Gln Arg Cys Ser Glu Phe Pro Cys Pro
               1025                1030                1035
Gln Trp Lys Ser Gly Asp Trp Ser Glu Cys Leu Val Thr Cys Gly
               1040                1045                1050
Lys Gly His Lys His Arg Gln Val Trp Cys Gln Phe Gly Glu Asp
               1055                1060                1065
Arg Leu Asn Asp Arg Met Cys Asp Pro Glu Thr Lys Pro Thr Ser
               1070                1075                1080
Met Gln Thr Cys Gln Gln Pro Glu Cys Ala Ser Trp Gln Ala Gly
               1085                1090                1095
Pro Trp Gly Gln Cys Ser Val Thr Cys Gly Gln Gly Tyr Gln Leu
               1100                1105                1110
Arg Ala Val Lys Cys Ile Ile Gly Thr Tyr Met Ser Val Val Asp
               1115                1120                1125
Asp Asn Asp Cys Asn Ala Ala Thr Arg Pro Thr Asp Thr Gln Asp
               1130                1135                1140
Cys Glu Leu Pro Ser Cys His Pro Pro Pro Ala Ala Pro Glu Thr
               1145                1150                1155
Arg Arg Ser Thr Tyr Ser Ala Pro Arg Thr Gln Trp Arg Phe Gly
```

```
                1160                1165                1170
Ser Trp Thr Pro Cys Ser Ala Thr Cys Gly Lys Gly Thr Arg Met
                1175                1180                1185
Arg Tyr Val Ser Cys Arg Asp Glu Asn Gly Ser Val Ala Asp Glu
                1190                1195                1200
Ser Ala Cys Ala Thr Leu Pro Arg Pro Val Ala Lys Glu Glu Cys
                1205                1210                1215
Ser Val Thr Pro Cys Gly Gln Trp Lys Ala Leu Asp Trp Ser Ser
                1220                1225                1230
Cys Ser Val Thr Cys Gly Gln Gly Arg Ala Thr Arg Gln Val Met
                1235                1240                1245
Cys Val Asn Tyr Ser Asp His Val Ile Asp Arg Ser Glu Cys Asp
                1250                1255                1260
Gln Asp Tyr Ile Pro Lys Thr Asp Gln Asp Cys Ser Met Ser Pro
                1265                1270                1275
Cys Pro Gln Arg Thr Pro Asp Ser Gly Leu Ala Gln His Pro Phe
                1280                1285                1290
Gln Asn Glu Asp Tyr Arg Pro Arg Ser Ala Ser Pro Ser Arg Thr
                1295                1300                1305
His Val Leu Gly Gly Asn Gln Trp Arg Thr Gly Pro Trp Gly Ala
                1310                1315                1320
Cys Ser Ser Thr Cys Ala Gly Gly Ser Gln Arg Arg Val Val Val
                1325                1330                1335
Cys Gln Asp Glu Asn Gly Tyr Thr Ala Asn Asp Cys Val Glu Arg
                1340                1345                1350
Ile Lys Pro Asp Glu Gln Arg Ala Cys Glu Ser Gly Pro Cys Pro
                1355                1360                1365
Gln Trp Ala Tyr Gly Asn Trp Gly Glu Cys Thr Lys Leu Cys Gly
                1370                1375                1380
Gly Gly Ile Arg Thr Arg Leu Val Val Cys Gln Arg Ser Asn Gly
                1385                1390                1395
Glu Arg Phe Pro Asp Leu Ser Cys Glu Ile Leu Asp Lys Pro Pro
                1400                1405                1410
Asp Arg Glu Gln Cys Asn Thr His Ala Cys Pro His Asp Ala Ala
                1415                1420                1425
Trp Ser Thr Gly Pro Trp Ser Ser Cys Ser Val Ser Cys Gly Arg
                1430                1435                1440
Gly His Lys Gln Arg Asn Val Tyr Cys Met Ala Lys Asp Gly Ser
                1445                1450                1455
His Leu Glu Ser Asp Tyr Cys Lys His Leu Ala Lys Pro His Gly
                1460                1465                1470
His Arg Lys Cys Arg Gly Gly Arg Cys Pro Lys Trp Lys Ala Gly
                1475                1480                1485
Ala Trp Ser Gln Cys Ser Val Ser Cys Gly Arg Gly Val Gln Gln
                1490                1495                1500
Arg His Val Gly Cys Gln Ile Gly Thr His Lys Ile Ala Arg Glu
                1505                1510                1515
Thr Glu Cys Asn Pro Tyr Thr Arg Pro Glu Ser Glu Arg Asp Cys
                1520                1525                1530
Gln Gly Pro Arg Cys Pro Leu Tyr Thr Trp Arg Ala Glu Glu Trp
                1535                1540                1545
Gln Glu Cys Thr Lys Thr Cys Gly Glu Gly Ser Arg Tyr Arg Lys
                1550                1555                1560
Val Val Cys Val Asp Asp Asn Lys Asn Glu Val His Gly Ala Arg
                1565                1570                1575
Cys Asp Val Ser Lys Arg Pro Val Asp Arg Glu Ser Cys Ser Leu
```

1580 1585 1590
Gln Pro Cys Glu Tyr Val Trp Ile Thr Gly Glu Trp Ser Glu Cys
1595 1600 1605
Ser Val Thr Cys Gly Lys Gly Tyr Lys Gln Arg Leu Val Ser Cys
1610 1615 1620
Ser Glu Ile Tyr Thr Gly Lys Glu Asn Tyr Glu Tyr Ser Tyr Gln
1625 1630 1635
Thr Thr Ile Asn Cys Pro Gly Thr Gln Pro Pro Ser Val His Pro
1640 1645 1650
Cys Tyr Leu Arg Asp Cys Pro Val Ser Ala Thr Trp Arg Val Gly
1655 1660 1665
Asn Trp Gly Ser Cys Ser Val Ser Cys Gly Val Gly Val Met Gln
1670 1675 1680
Arg Ser Val Gln Cys Leu Thr Asn Glu Asp Gln Pro Ser His Leu
1685 1690 1695
Cys His Thr Asp Leu Lys Pro Glu Glu Arg Lys Thr Cys Arg Asn
1700 1705 1710
Val Tyr Asn Cys Glu Leu Pro Gln Asn Cys Lys Glu Val Lys Arg
1715 1720 1725
Leu Lys Gly Ala Ser Glu Asp Gly Glu Tyr Phe Leu Met Ile Arg
1730 1735 1740
Gly Lys Leu Leu Lys Ile Phe Cys Ala Gly Met His Ser Asp His
1745 1750 1755
Pro Lys Glu Tyr Val Thr Leu Val His Gly Asp Ser Glu Asn Phe
1760 1765 1770
Ser Glu Val Tyr Gly His Arg Leu His Asn Pro Thr Glu Cys Pro
1775 1780 1785
Tyr Asn Gly Ser Arg Arg Asp Asp Cys Gln Cys Arg Lys Asp Tyr
1790 1795 1800
Thr Ala Ala Gly Phe Ser Ser Phe Gln Lys Ile Arg Ile Asp Leu
1805 1810 1815
Thr Ser Met Gln Ile Ile Thr Thr Asp Leu Gln Phe Ala Arg Thr
1820 1825 1830
Ser Glu Gly His Pro Val Pro Phe Ala Thr Ala Gly Asp Cys Tyr
1835 1840 1845
Ser Ala Ala Lys Cys Pro Gln Gly Arg Phe Ser Ile Asn Leu Tyr
1850 1855 1860
Gly Thr Gly Leu Ser Leu Thr Glu Ser Ala Arg Trp Ile Ser Gln
1865 1870 1875
Gly Asn Tyr Ala Val Ser Asp Ile Lys Lys Ser Pro Asp Gly Thr
1880 1885 1890
Arg Val Val Gly Lys Cys Gly Gly Tyr Cys Gly Lys Cys Thr Pro
1895 1900 1905
Ser Ser Gly Thr Gly Leu Glu Val Arg Val Leu
1910 1915

<210> 11
<211> 314
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3965293CD1

<400> 11

```
Met Glu Asp Asp Ser Leu Tyr Leu Gly Gly Glu Trp Gln Phe Asn
  1               5                  10                  15
His Phe Ser Lys Leu Thr Ser Ser Arg Pro Asp Ala Ala Phe Ala
             20                  25                  30
Glu Ile Gln Arg Thr Ser Leu Pro Glu Lys Ser Pro Leu Ser Cys
             35                  40                  45
Glu Thr Arg Val Asp Leu Cys Asp Asp Leu Ala Pro Val Ala Arg
             50                  55                  60
Gln Leu Ala Pro Arg Glu Lys Leu Pro Leu Ser Ser Arg Arg Pro
             65                  70                  75
Ala Ala Val Gly Ala Gly Leu Gln Asn Met Gly Asn Thr Cys Tyr
             80                  85                  90
Val Asn Ala Ser Leu Gln Cys Leu Thr Tyr Thr Pro Pro Leu Ala
             95                 100                 105
Asn Tyr Met Leu Ser Arg Glu His Ser Gln Thr Cys His Arg His
            110                 115                 120
Lys Gly Cys Met Leu Cys Thr Met Gln Ala His Ile Thr Arg Ala
            125                 130                 135
Leu His Asn Pro Gly His Val Ile Gln Pro Ser Gln Ala Leu Ala
            140                 145                 150
Ala Gly Phe His Arg Gly Lys Gln Glu Asp Ala His Glu Phe Leu
            155                 160                 165
Met Phe Thr Val Asp Ala Met Lys Lys Ala Cys Leu Pro Gly His
            170                 175                 180
Lys Gln Val Asp His His Ser Lys Asp Thr Thr Leu Ile His Gln
            185                 190                 195
Ile Phe Gly Gly Tyr Trp Arg Ser Gln Ile Lys Cys Leu His Cys
            200                 205                 210
His Gly Ile Ser Asp Thr Phe Asp Pro Tyr Leu Asp Ile Ala Leu
            215                 220                 225
Asp Ile Gln Ala Ala Gln Ser Val Gln Gln Ala Leu Glu Gln Leu
            230                 235                 240
Val Lys Pro Glu Glu Leu Asn Gly Glu Asn Ala Tyr His Cys Gly
            245                 250                 255
Val Cys Leu Gln Arg Ala Pro Ala Ser Lys Thr Leu Thr Leu His
            260                 265                 270
Thr Ser Ala Lys Val Leu Ile Leu Val Leu Lys Arg Phe Ser Asp
            275                 280                 285
Val Thr Gly Asn Leu Glu Pro Asn Ser Ala Arg Ala Arg Ala Glu
            290                 295                 300
Arg Ser Gln Cys Ser Thr Ser Pro Cys Pro Ser Cys Arg Gly
            305                 310
```

<210> 12
<211> 437
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4948403CD1

<400> 12

```
Met Lys Cys Leu Gly Lys Arg Arg Gly Gln Ala Ala Ala Phe Leu
  1               5                  10                  15
Pro Leu Cys Trp Leu Phe Leu Lys Ile Leu Gln Pro Gly His Ser
```

```
                20                  25                  30
His Leu Tyr Asn Asn Arg Tyr Ala Gly Asp Lys Val Ile Arg Phe
                35                  40                  45
Ile Pro Lys Thr Glu Glu Glu Ala Tyr Ala Leu Lys Lys Ile Ser
                50                  55                  60
Tyr Gln Leu Lys Val Asp Leu Trp Gln Pro Ser Ser Ile Ser Tyr
                65                  70                  75
Val Ser Glu Gly Thr Val Thr Asp Val His Ile Pro Gln Asn Gly
                80                  85                  90
Ser Arg Ala Leu Leu Ala Phe Leu Gln Glu Ala Asn Ile Gln Tyr
                95                  100                 105
Lys Val Leu Ile Glu Asp Leu Gln Lys Thr Leu Glu Lys Gly Ser
                110                 115                 120
Ser Leu His Thr Gln Arg Asn Arg Arg Ser Leu Ser Gly Tyr Asn
                125                 130                 135
Tyr Glu Val Tyr His Ser Leu Glu Glu Ile Gln Asn Trp Met His
                140                 145                 150
His Leu Asn Lys Thr His Ser Gly Leu Ile His Met Phe Ser Ile
                155                 160                 165
Gly Arg Ser Tyr Glu Gly Arg Ser Leu Phe Ile Leu Lys Leu Gly
                170                 175                 180
Arg Arg Ser Arg Leu Lys Arg Ala Val Trp Ile Asp Cys Gly Ile
                185                 190                 195
His Ala Arg Glu Trp Ile Gly Pro Ala Phe Cys Gln Trp Phe Val
                200                 205                 210
Lys Glu Ala Leu Leu Thr Tyr Lys Ser Asp Pro Ala Met Arg Lys
                215                 220                 225
Met Leu Asn His Leu Tyr Phe Tyr Ile Met Pro Val Phe Asn Val
                230                 235                 240
Asp Gly Tyr His Phe Ser Trp Thr Asn Asp Arg Phe Trp Arg Lys
                245                 250                 255
Thr Arg Ser Arg Asn Ser Arg Phe Arg Cys Arg Gly Val Asp Ala
                260                 265                 270
Asn Arg Asn Trp Lys Val Lys Trp Cys Asp Glu Gly Ala Ser Met
                275                 280                 285
His Pro Cys Asp Asp Thr Tyr Cys Gly Pro Phe Pro Glu Ser Glu
                290                 295                 300
Pro Glu Val Lys Ala Val Ala Asn Phe Leu Arg Lys His Arg Lys
                305                 310                 315
His Ile Arg Ala Tyr Leu Ser Phe His Ala Tyr Ala Gln Met Leu
                320                 325                 330
Leu Tyr Pro Tyr Ser Tyr Lys Tyr Ala Thr Ile Pro Asn Phe Arg
                335                 340                 345
Cys Val Glu Ser Ala Ala Tyr Lys Ala Val Asn Ala Leu Gln Ser
                350                 355                 360
Val Tyr Gly Val Arg Tyr Arg Tyr Gly Pro Ala Ser Thr Thr Leu
                365                 370                 375
Tyr Val Ser Ser Gly Ser Ser Met Asp Trp Ala Tyr Lys Asn Gly
                380                 385                 390
Ile Pro Tyr Ala Phe Ala Phe Glu Leu Arg Asp Thr Gly Tyr Phe
                395                 400                 405
Gly Phe Leu Leu Pro Glu Met Leu Ile Lys Pro Thr Cys Thr Glu
                410                 415                 420
Thr Met Leu Ala Val Lys Asn Ile Thr Met His Leu Leu Lys Lys
                425                 430                 435
Cys Pro
```

<210> 13
<211> 742
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473165CD1

<400> 13

```
Met Val Glu Ser Ala Gly Arg Ala Gly Gln Lys Arg Pro Gly Phe
  1               5                  10                  15
Leu Glu Gly Gly Leu Leu Leu Leu Leu Leu Leu Val Thr Ala Ala
                 20                  25                  30
Leu Val Ala Leu Gly Val Leu Tyr Ala Asp Arg Arg Gly Ile Pro
                 35                  40                  45
Glu Ala Gln Glu Val Ser Glu Val Cys Thr Thr Pro Gly Cys Val
                 50                  55                  60
Ile Ala Ala Ala Arg Ile Leu Gln Asn Met Asp Pro Thr Thr Glu
                 65                  70                  75
Pro Cys Asp Asp Phe Tyr Gln Phe Ala Cys Gly Gly Trp Leu Arg
                 80                  85                  90
Arg His Val Ile Pro Glu Thr Asn Ser Arg Tyr Ser Ile Phe Asp
                 95                 100                 105
Val Leu Arg Asp Glu Leu Glu Val Ile Leu Lys Ala Val Leu Glu
                110                 115                 120
Asn Ser Thr Ala Lys Asp Arg Pro Ala Val Glu Lys Ala Arg Thr
                125                 130                 135
Leu Tyr Arg Ser Cys Met Asn Gln Ser Val Ile Glu Lys Arg Gly
                140                 145                 150
Ser Gln Pro Leu Leu Asp Ile Leu Glu Val Val Gly Gly Trp Pro
                155                 160                 165
Val Ala Met Asp Arg Trp Asn Glu Thr Val Gly Leu Glu Trp Glu
                170                 175                 180
Leu Glu Arg Gln Leu Ala Leu Met Asn Ser Gln Phe Asn Arg Arg
                185                 190                 195
Val Leu Ile Asp Leu Phe Ile Trp Asn Asp Asp Gln Asn Ser Ser
                200                 205                 210
Arg His Ile Ile Tyr Ile Asp Gln Pro Thr Leu Gly Met Pro Ser
                215                 220                 225
Arg Glu Tyr Tyr Phe Asn Gly Gly Ser Asn Arg Lys Val Arg Glu
                230                 235                 240
Ala Tyr Leu Gln Phe Met Val Ser Val Ala Thr Leu Leu Arg Glu
                245                 250                 255
Asp Ala Asn Leu Pro Arg Asp Ser Cys Leu Val Gln Glu Asp Met
                260                 265                 270
Val Gln Val Leu Glu Leu Glu Thr Gln Leu Ala Lys Ala Thr Val
                275                 280                 285
Pro Gln Glu Glu Arg His Asp Val Ile Ala Leu Tyr His Arg Met
                290                 295                 300
Gly Leu Glu Glu Leu Gln Ser Gln Phe Gly Leu Lys Gly Phe Asn
                305                 310                 315
Trp Thr Leu Phe Ile Gln Thr Val Leu Ser Ser Val Lys Ile Lys
                320                 325                 330
```

```
Leu Leu Pro Asp Glu Glu Val Val Val Tyr Gly Ile Pro Tyr Leu
                335                 340                 345
Gln Asn Leu Glu Asn Ile Ile Asp Thr Tyr Ser Ala Arg Thr Ile
                350                 355                 360
Gln Asn Tyr Leu Val Trp Arg Leu Val Leu Asp Arg Ile Gly Ser
                365                 370                 375
Leu Ser Gln Arg Phe Lys Asp Thr Arg Val Asn Tyr Arg Lys Ala
                380                 385                 390
Leu Phe Gly Thr Met Val Glu Glu Val Arg Trp Arg Glu Cys Val
                395                 400                 405
Gly Tyr Val Asn Ser Asn Met Glu Asn Ala Val Gly Ser Leu Tyr
                410                 415                 420
Val Arg Glu Ala Phe Pro Gly Asp Ser Lys Ser Met Val Glu Leu
                425                 430                 435
Ile Asp Lys Val Arg Thr Val Phe Val Glu Thr Leu Asp Glu Leu
                440                 445                 450
Gly Trp Met Asp Glu Glu Ser Lys Lys Lys Ala Gln Glu Lys Ala
                455                 460                 465
Met Ser Ile Arg Glu Gln Ile Gly His Pro Asp Tyr Ile Leu Glu
                470                 475                 480
Glu Met Asn Arg Arg Leu Asp Glu Glu Tyr Ser Asn Val Asn Phe
                485                 490                 495
Ser Glu Asp Leu Tyr Phe Glu Asn Ser Leu Gln Asn Leu Lys Val
                500                 505                 510
Gly Ala Gln Arg Ser Leu Arg Lys Leu Arg Glu Lys Val Asp Pro
                515                 520                 525
Asn Leu Ile Ile Gly Ala Ala Val Val Asn Ala Phe Tyr Ser Pro
                530                 535                 540
Asn Arg Asn Gln Ile Val Phe Pro Ala Gly Ile Leu Gln Pro Pro
                545                 550                 555
Phe Phe Ser Lys Glu Gln Pro Gln Ala Leu Asn Phe Gly Gly Ile
                560                 565                 570
Gly Met Val Ile Gly His Glu Ile Thr His Gly Phe Asp Asp Asn
                575                 580                 585
Gly Arg Asn Phe Asp Lys Asn Gly Asn Met Met Asp Trp Trp Ser
                590                 595                 600
Asn Phe Ser Thr Gln His Phe Arg Glu Gln Ser Glu Cys Met Ile
                605                 610                 615
Tyr Gln Tyr Gly Asn Tyr Ser Trp Asp Leu Ala Asp Glu Gln Asn
                620                 625                 630
Val Asn Gly Phe Asn Thr Leu Gly Glu Asn Ile Ala Asp Asn Gly
                635                 640                 645
Gly Val Arg Gln Ala Tyr Lys Ala Tyr Leu Lys Trp Met Ala Glu
                650                 655                 660
Gly Gly Lys Asp Gln Gln Leu Pro Gly Leu Asp Leu Thr His Glu
                665                 670                 675
Gln Leu Phe Phe Ile Asn Tyr Ala Gln Val Trp Cys Gly Ser Tyr
                680                 685                 690
Arg Pro Glu Phe Ala Ile Gln Ser Ile Lys Thr Asp Val His Ser
                695                 700                 705
Pro Leu Lys Tyr Arg Val Leu Gly Ser Leu Gln Asn Leu Ala Ala
                710                 715                 720
Phe Ala Asp Thr Phe His Cys Ala Arg Gly Thr Pro Met His Pro
                725                 730                 735
Lys Glu Arg Cys Arg Val Trp
                740
```

<210> 14
<211> 582
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476667CD1

<400> 14

```
Met Phe Thr Leu Thr Thr Asn Gly Asp Leu Pro Arg Pro Ile Phe
  1               5                  10                  15
Ile Pro Asn Gly Met Pro Asn Thr Val Val Pro Cys Gly Thr Glu
                20                  25                  30
Lys Asn Phe Thr Asn Gly Met Val Asn Gly His Met Pro Ser Leu
                35                  40                  45
Pro Asp Ser Pro Phe Thr Gly Tyr Ile Ile Ala Val His Arg Lys
                50                  55                  60
Met Met Arg Thr Glu Leu Tyr Phe Leu Ser Ser Gln Lys Asn Arg
                65                  70                  75
Pro Ser Leu Phe Gly Met Pro Leu Ile Val Pro Cys Thr Val His
                80                  85                  90
Thr Arg Lys Lys Asp Leu Tyr Asp Ala Val Trp Ile Gln Val Ser
                95                 100                 105
Arg Leu Ala Ser Pro Leu Pro Pro Gln Glu Ala Ser Asn His Ala
               110                 115                 120
Gln Asp Cys Asp Asp Ser Met Gly Tyr Gln Tyr Pro Phe Thr Leu
               125                 130                 135
Arg Val Val Gln Lys Asp Gly Asn Ser Cys Ala Trp Cys Pro Trp
               140                 145                 150
Tyr Arg Phe Cys Arg Gly Cys Lys Ile Asp Cys Gly Glu Asp Arg
               155                 160                 165
Ala Phe Ile Gly Asn Ala Tyr Ile Ala Val Asp Trp Asp Pro Thr
               170                 175                 180
Ala Leu His Leu Arg Tyr Gln Thr Ser Gln Glu Arg Val Val Asp
               185                 190                 195
Glu His Glu Ser Val Glu Gln Ser Arg Arg Ala Gln Ala Glu Pro
               200                 205                 210
Ile Asn Leu Asp Ser Cys Leu Arg Ala Phe Thr Ser Glu Glu Glu
               215                 220                 225
Leu Gly Glu Asn Glu Met Tyr Tyr Cys Ser Lys Cys Lys Thr His
               230                 235                 240
Cys Leu Ala Thr Lys Lys Leu Asp Leu Trp Arg Leu Pro Pro Ile
               245                 250                 255
Leu Ile Ile His Leu Lys Arg Phe Gln Phe Val Asn Gly Arg Trp
               260                 265                 270
Ile Lys Ser Gln Lys Ile Val Lys Phe Pro Arg Glu Ser Phe Asp
               275                 280                 285
Pro Ser Ala Phe Leu Val Pro Arg Asp Pro Ala Leu Cys Gln His
               290                 295                 300
Lys Pro Leu Thr Pro Gln Gly Asp Glu Leu Ser Glu Pro Arg Ile
               305                 310                 315
Leu Ala Arg Glu Val Lys Lys Val Asp Ala Gln Ser Ser Ala Gly
               320                 325                 330
Glu Glu Asp Val Leu Leu Ser Lys Ser Pro Ser Ser Leu Ser Ala
               335                 340                 345
```

```
Asn Ile Ile Ser Ser Pro Lys Gly Ser Pro Ser Ser Ser Arg Lys
                350                 355                 360
Ser Gly Thr Ser Cys Pro Ser Ser Lys Asn Ser Ser Pro Asn Ser
                365                 370                 375
Ser Pro Arg Thr Leu Gly Arg Ser Lys Gly Arg Leu Arg Leu Pro
                380                 385                 390
Gln Ile Gly Ser Lys Asn Lys Leu Ser Ser Ser Lys Glu Asn Leu
                395                 400                 405
Asp Ala Ser Lys Glu Asn Gly Ala Gly Gln Ile Cys Glu Leu Ala
                410                 415                 420
Asp Ala Leu Ser Arg Gly His Val Leu Gly Gly Ser Gln Pro Glu
                425                 430                 435
Leu Val Thr Pro Gln Asp His Glu Val Ala Leu Ala Asn Gly Phe
                440                 445                 450
Leu Tyr Glu His Glu Ala Cys Gly Asn Gly Tyr Ser Asn Gly Gln
                455                 460                 465
Leu Gly Asn His Ser Glu Glu Asp Ser Thr Asp Asp Gln Arg Glu
                470                 475                 480
Asp Thr Arg Ile Lys Pro Ile Tyr Asn Leu Tyr Ala Ile Ser Cys
                485                 490                 495
His Ser Gly Ile Leu Gly Gly Gly His Tyr Val Thr Tyr Ala Lys
                500                 505                 510
Asn Pro Asn Cys Lys Trp Tyr Cys Tyr Asn Asp Ser Ser Cys Lys
                515                 520                 525
Glu Leu His Pro Asp Glu Ile Asp Thr Asp Ser Ala Tyr Ile Leu
                530                 535                 540
Phe Tyr Glu Gln Gln Gly Ile Asp Tyr Ala Gln Phe Leu Pro Lys
                545                 550                 555
Thr Asp Gly Lys Lys Met Ala Asp Thr Ser Ser Met Asp Glu Asp
                560                 565                 570
Phe Glu Ser Asp Tyr Lys Lys Tyr Cys Val Leu Gln
                575                 580
```

<210> 15
<211> 290
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7479166CD1

<400> 15

```
Met Leu Ser Pro Pro Gln Pro Arg Thr Pro Asp Cys Arg Leu Gln
  1               5                  10                  15
Ala Ser Leu Glu Ala Leu Ala Thr Leu Ala Pro Gln Pro Ser Asp
                20                  25                  30
Trp Leu Cys Phe Ala Asp Leu Gly Trp Phe Glu Ala Asp Gly Ala
                35                  40                  45
Ala His Ser Met Gly Leu Gly Ser Ser Leu Lys Trp Ala Trp Ala
                50                  55                  60
Lys Pro Ser Gly Met Pro Val Pro Glu Asn Asp Leu Val Gly Ile
                65                  70                  75
Val Gly Gly His Asn Ala Pro Pro Gly Lys Trp Pro Trp Gln Val
                80                  85                  90
Ser Leu Arg Val Tyr Ser Tyr His Trp Ala Ser Trp Ala His Ile
```

```
                95                  100                 105
Cys Gly Gly Ser Leu Ile His Pro Gln Trp Val Leu Thr Ala Ala
                110                 115                 120
His Cys Ile Phe Trp Lys Asp Thr Asp Pro Ser Ile Tyr Arg Ile
                125                 130                 135
His Ala Gly Asp Val Tyr Leu Tyr Gly Gly Arg Gly Leu Leu Asn
                140                 145                 150
Val Ser Arg Ile Ile Val His Pro Asn Tyr Val Thr Ala Gly Leu
                155                 160                 165
Gly Ala Asp Val Ala Leu Leu Gln Leu Pro Gly Ser Pro Leu Ser
                170                 175                 180
Pro Glu Ser Leu Pro Pro Pro Tyr Arg Leu Gln Gln Ala Ser Val
                185                 190                 195
Gln Val Leu Glu Asn Ala Val Cys Glu Gln Pro Tyr Arg Asn Ala
                200                 205                 210
Ser Gly His Thr Gly Asp Arg Gln Leu Ile Leu Asp Asp Met Leu
                215                 220                 225
Cys Ala Gly Ser Glu Gly Arg Asp Ser Cys Tyr Gly Asp Ser Gly
                230                 235                 240
Gly Pro Leu Val Cys Arg Leu Arg Gly Ser Trp Arg Leu Val Gly
                245                 250                 255
Val Val Ser Trp Gly Tyr Gly Cys Thr Leu Arg Asp Phe Pro Gly
                260                 265                 270
Val Tyr Thr His Val Gln Ile Tyr Val Leu Trp Ile Leu Gln Gln
                275                 280                 285
Val Gly Glu Leu Pro
                290
```

<210> 16
<211> 708
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3671788CD1

<400> 16

```
Met Ala Ser Ser Ser Gly Arg Val Thr Ile Gln Leu Val Asp Glu
  1                 5                  10                  15
Glu Ala Gly Val Gly Ala Gly Arg Leu Gln Leu Phe Arg Gly Gln
                20                  25                  30
Ser Tyr Glu Ala Ile Arg Ala Ala Cys Leu Asp Ser Gly Ile Leu
                35                  40                  45
Phe Arg Asp Pro Tyr Phe Pro Ala Gly Pro Asp Ala Leu Gly Tyr
                50                  55                  60
Asp Gln Leu Gly Pro Asp Ser Glu Lys Ala Lys Gly Val Lys Trp
                65                  70                  75
Met Arg Pro His Glu Phe Cys Ala Glu Pro Lys Phe Ile Cys Glu
                80                  85                  90
Asp Met Ser Arg Thr Asp Val Cys Gln Gly Ser Leu Gly Asn Cys
                95                  100                 105
Trp Phe Leu Ala Ala Ala Ala Ser Leu Thr Leu Tyr Pro Arg Leu
                110                 115                 120
Leu Arg Arg Val Val Pro Pro Gly Gln Asp Phe Gln His Gly Tyr
                125                 130                 135
```

```
Ala Gly Val Phe His Phe Gln Leu Trp Gln Phe Gly Arg Trp Met
                140                 145                 150
Asp Val Val Val Asp Asp Arg Leu Pro Val Arg Glu Gly Lys Leu
            155                 160                 165
Met Phe Val Arg Ser Glu Gln Arg Asn Glu Phe Trp Ala Pro Leu
            170                 175                 180
Leu Glu Lys Ala Tyr Ala Lys Leu His Gly Ser Tyr Glu Val Met
            185                 190                 195
Arg Gly Gly His Met Asn Glu Ala Phe Val Asp Phe Thr Gly Gly
            200                 205                 210
Val Gly Glu Val Leu Tyr Leu Arg Gln Asn Ser Met Gly Leu Phe
            215                 220                 225
Ser Ala Leu Arg His Ala Leu Ala Lys Glu Ser Leu Val Gly Ala
            230                 235                 240
Thr Ala Leu Ser Asp Arg Gly Glu Tyr Arg Thr Glu Glu Gly Leu
            245                 250                 255
Val Lys Gly His Ala Tyr Ser Ile Thr Gly Thr His Lys Val Phe
            260                 265                 270
Leu Gly Phe Thr Lys Val Arg Leu Leu Arg Leu Arg Asn Pro Trp
            275                 280                 285
Gly Cys Val Glu Trp Thr Gly Ala Trp Ser Asp Ser Cys Pro Arg
            290                 295                 300
Trp Asp Thr Leu Pro Thr Glu Cys Arg Asp Ala Leu Leu Val Lys
            305                 310                 315
Lys Glu Asp Gly Glu Phe Trp Met Glu Leu Arg Asp Phe Leu Leu
            320                 325                 330
His Phe Asp Thr Val Gln Ile Cys Ser Leu Ser Pro Glu Val Leu
            335                 340                 345
Gly Pro Ser Pro Glu Gly Gly Gly Trp His Val His Thr Phe Gln
            350                 355                 360
Gly Arg Trp Val Arg Gly Phe Asn Ser Gly Gly Ser Gln Pro Asn
            365                 370                 375
Ala Glu Thr Phe Trp Thr Asn Pro Gln Phe Arg Leu Thr Leu Leu
            380                 385                 390
Glu Pro Asp Glu Glu Asp Asp Glu Asp Glu Glu Gly Pro Trp Gly
            395                 400                 405
Gly Trp Gly Ala Ala Gly Ala Arg Gly Pro Ala Arg Gly Gly Arg
            410                 415                 420
Thr Pro Lys Cys Thr Val Leu Leu Ser Leu Ile Gln Arg Asn Arg
            425                 430                 435
Arg Arg Leu Arg Ala Lys Gly Leu Thr Tyr Leu Thr Val Gly Phe
            440                 445                 450
His Val Phe Gln Ala Glu Gly Ser Thr Gly Thr Asp Asn Glu Arg
            455                 460                 465
Thr His Gly Phe Thr Gly His Arg Gly Ala Gln Leu Ala Gly His
            470                 475                 480
Thr His Gly Pro Gln Glu Ala Ser Lys Arg Tyr Thr Gln Asn Ser
            485                 490                 495
Ala Glu Val Ala Pro Asp Arg Glu Ala Asp Asp Asp Gly Gly Gln
            500                 505                 510
Gly Phe Gly Asp Gly Pro Trp Glu Ile Asp Asp Val Ile Ser Ala
            515                 520                 525
Asp Leu Gln Ser Leu Gln Gly Pro Tyr Leu Pro Leu Glu Leu Gly
            530                 535                 540
Leu Glu Gln Leu Phe Gln Glu Leu Ala Gly Glu Glu Glu Glu Leu
            545                 550                 555
```

```
Asn Ala Ser Gln Leu Gln Ala Leu Leu Ser Ile Ala Leu Glu Pro
                560                 565                 570
Ala Arg Ala His Thr Ser Thr Pro Arg Glu Ile Gly Leu Arg Thr
                575                 580                 585
Cys Glu Gln Leu Leu Gln Cys Phe Gly His Gly Gln Ser Leu Ala
                590                 595                 600
Leu His His Phe Gln Gln Leu Trp Gly Tyr Leu Leu Glu Trp Gln
                605                 610                 615
Ala Ile Phe Asn Lys Phe Asp Glu Asp Thr Ser Gly Thr Met Asn
                620                 625                 630
Ser Tyr Glu Leu Arg Leu Ala Leu Asn Ala Ala Gly Phe His Leu
                635                 640                 645
Asn Asn Gln Leu Thr Gln Thr Leu Thr Ser Arg Tyr Arg Asp Ser
                650                 655                 660
Arg Leu Arg Val Asp Phe Glu Arg Phe Val Ser Cys Val Ala His
                665                 670                 675
Leu Thr Cys Ile Phe Cys His Cys Ser Gln His Leu Asp Gly Gly
                680                 685                 690
Glu Gly Val Ile Cys Leu Thr His Arg Gln Trp Met Glu Val Ala
                695                 700                 705
Thr Phe Ser
```

<210> 17
<211> 649
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7479181CD1

<400> 17

```
Met Glu Leu Gly Cys Trp Thr Gln Leu Gly Leu Thr Phe Leu Gln
  1                 5                  10                  15
Leu Leu Leu Ile Ser Ser Leu Pro Arg Glu Tyr Thr Val Ile Asn
                 20                  25                  30
Glu Ala Cys Pro Gly Ala Glu Trp Asn Ile Met Cys Arg Glu Cys
                 35                  40                  45
Cys Glu Tyr Asp Gln Ile Glu Cys Val Cys Pro Gly Lys Arg Glu
                 50                  55                  60
Val Val Gly Tyr Thr Ile Pro Cys Cys Arg Asn Glu Glu Asn Glu
                 65                  70                  75
Cys Asp Ser Cys Leu Ile His Pro Gly Cys Thr Ile Phe Glu Asn
                 80                  85                  90
Cys Lys Ser Cys Arg Asn Gly Ser Trp Gly Gly Thr Leu Asp Asp
                 95                 100                 105
Phe Tyr Val Lys Gly Phe Tyr Cys Ala Glu Cys Arg Ala Gly Trp
                110                 115                 120
Tyr Gly Gly Asp Cys Met Arg Cys Gly Gln Val Leu Arg Ala Pro
                125                 130                 135
Lys Gly Gln Ile Leu Leu Glu Ser Tyr Pro Leu Asn Ala His Cys
                140                 145                 150
Glu Trp Thr Ile His Ala Lys Pro Gly Phe Val Ile Gln Leu Arg
                155                 160                 165
Phe Val Met Leu Ser Leu Glu Phe Asp Tyr Met Cys Gln Tyr Asp
```

```
                170                 175                 180
Tyr Val Glu Val Arg Asp Gly Asp Asn Arg Asp Gly Gln Ile Ile
                185                 190                 195
Lys Arg Val Cys Gly Asn Glu Arg Pro Ala Pro Ile Gln Ser Ile
                200                 205                 210
Gly Ser Ser Leu His Val Leu Phe His Ser Asp Gly Ser Lys Asn
                215                 220                 225
Phe Asp Gly Phe His Ala Ile Tyr Glu Glu Ile Thr Ala Cys Ser
                230                 235                 240
Ser Ser Pro Cys Phe His Asp Gly Thr Cys Val Leu Asp Lys Ala
                245                 250                 255
Gly Ser Tyr Lys Cys Ala Cys Leu Ala Gly Tyr Thr Gly Gln Arg
                260                 265                 270
Cys Glu Asn Pro Cys Arg Glu Pro Lys Ile Ser Asp Leu Val Arg
                275                 280                 285
Arg Arg Val Leu Pro Met Gln Val Gln Ser Arg Glu Thr Pro Leu
                290                 295                 300
His Gln Leu Tyr Ser Ala Ala Phe Ser Lys Gln Lys Leu Gln Ser
                305                 310                 315
Ala Pro Thr Lys Lys Pro Ala Leu Pro Phe Gly Asp Leu Pro Met
                320                 325                 330
Gly Tyr Gln His Leu His Thr Gln Leu Gln Tyr Glu Cys Ile Ser
                335                 340                 345
Pro Phe Tyr Arg Arg Leu Gly Ser Ser Arg Arg Thr Cys Leu Arg
                350                 355                 360
Thr Gly Lys Trp Ser Gly Arg Ala Pro Ser Cys Ile Pro Ile Cys
                365                 370                 375
Gly Lys Ile Glu Asn Ile Thr Ala Pro Lys Thr Gln Gly Leu Arg
                380                 385                 390
Trp Pro Trp Gln Ala Ala Ile Tyr Arg Arg Thr Ser Gly Val His
                395                 400                 405
Asp Gly Ser Leu His Lys Gly Ala Trp Phe Leu Val Cys Ser Gly
                410                 415                 420
Ala Leu Val Asn Glu Arg Thr Val Val Val Ala Ala His Cys Val
                425                 430                 435
Thr Asp Leu Gly Lys Val Thr Met Ile Lys Thr Ala Asp Leu Lys
                440                 445                 450
Val Val Leu Gly Lys Phe Tyr Arg Asp Asp Asp Arg Asp Glu Lys
                455                 460                 465
Thr Ile Gln Ser Leu Gln Ile Ser Ala Ile Ile Leu His Pro Asn
                470                 475                 480
Tyr Asp Pro Ile Leu Leu Asp Ala Asp Ile Ala Ile Leu Lys Leu
                485                 490                 495
Leu Asp Lys Ala Arg Ile Ser Thr Arg Val Gln Pro Ile Cys Leu
                500                 505                 510
Ala Ala Ser Arg Asp Leu Ser Thr Ser Phe Gln Glu Ser His Ile
                515                 520                 525
Thr Val Ala Gly Trp Asn Val Leu Ala Asp Val Arg Ser Pro Gly
                530                 535                 540
Phe Lys Asn Asp Thr Leu Arg Ser Gly Val Val Ser Val Val Asp
                545                 550                 555
Ser Leu Leu Cys Glu Glu Gln His Glu Asp His Gly Ile Pro Val
                560                 565                 570
Ser Val Thr Asp Asn Met Phe Cys Ala Ser Trp Glu Pro Thr Ala
                575                 580                 585
Pro Ser Asp Ile Cys Thr Ala Glu Thr Gly Gly Ile Ala Ala Val
```

590 595 600
Ser Phe Pro Gly Arg Ala Ser Pro Glu Pro Arg Trp His Leu Met
605 610 615
Gly Leu Val Ser Trp Ser Tyr Asp Lys Thr Cys Ser His Arg Leu
620 625 630
Ser Thr Ala Phe Thr Lys Val Leu Pro Phe Lys Asp Trp Ile Glu
635 640 645
Arg Asn Met Lys

<210> 18
<211> 918
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6621372CD1

<400> 18
Met Pro Gly Gly Ala Gly Ala Ala Arg Leu Cys Leu Leu Ala Phe
1 5 10 15
Ala Leu Gln Pro Leu Arg Pro Arg Ala Ala Arg Glu Pro Gly Trp
20 25 30
Thr Arg Gly Ser Glu Glu Gly Ser Pro Lys Leu Gln His Glu Leu
35 40 45
Ile Ile Pro Gln Trp Lys Thr Ser Glu Ser Pro Val Arg Glu Lys
50 55 60
His Pro Leu Lys Ala Glu Leu Arg Val Met Ala Glu Gly Arg Glu
65 70 75
Leu Ile Leu Asp Leu Glu Lys Asn Glu Gln Leu Phe Ala Pro Ser
80 85 90
Tyr Thr Glu Thr His Tyr Thr Ser Ser Gly Asn Pro Gln Thr Thr
95 100 105
Thr Arg Lys Leu Glu Asp His Cys Phe Tyr His Gly Thr Val Arg
110 115 120
Glu Thr Glu Leu Ser Ser Val Thr Leu Ser Thr Cys Arg Gly Ile
125 130 135
Arg Gly Leu Ile Thr Val Ser Ser Asn Leu Ser Tyr Val Ile Glu
140 145 150
Pro Leu Pro Asp Ser Lys Gly Gln His Leu Ile Tyr Arg Ser Glu
155 160 165
His Leu Lys Pro Pro Pro Gly Asn Cys Gly Phe Glu His Ser Lys
170 175 180
Pro Thr Thr Arg Asp Trp Ala Leu Gln Phe Thr Gln Gln Thr Lys
185 190 195
Lys Arg Pro Arg Arg Met Lys Arg Glu Asp Leu Asn Ser Met Lys
200 205 210
Tyr Val Glu Leu Tyr Leu Val Ala Asp Tyr Leu Glu Phe Gln Lys
215 220 225
Asn Arg Arg Asp Gln Asp Ala Thr Lys His Lys Leu Ile Glu Ile
230 235 240
Ala Asn Tyr Val Asp Lys Phe Tyr Arg Ser Leu Asn Ile Arg Ile
245 250 255
Ala Leu Val Gly Leu Glu Val Trp Thr His Gly Asn Met Cys Glu
260 265 270

```
Val Ser Glu Asn Pro Tyr Ser Thr Leu Trp Ser Phe Leu Ser Trp
                275                 280                 285
Arg Arg Lys Leu Leu Ala Gln Lys Tyr His Asp Asn Ala Gln Leu
                290                 295                 300
Ile Thr Gly Met Ser Phe His Gly Thr Thr Ile Gly Leu Ala Pro
                305                 310                 315
Leu Met Ala Met Cys Ser Val Tyr Gln Ser Gly Gly Val Asn Met
                320                 325                 330
Asp His Ser Glu Asn Ala Ile Gly Val Ala Ala Thr Met Ala His
                335                 340                 345
Glu Met Gly His Asn Phe Gly Met Thr His Asp Ser Ala Asp Cys
                350                 355                 360
Cys Ser Ala Ser Ala Ala Asp Gly Gly Cys Ile Met Ala Ala Ala
                365                 370                 375
Thr Gly His Pro Phe Pro Lys Val Phe Asn Gly Cys Asn Arg Arg
                380                 385                 390
Glu Leu Asp Arg Tyr Leu Gln Ser Gly Gly Gly Met Cys Leu Ser
                395                 400                 405
Asn Met Pro Asp Thr Arg Met Leu Tyr Gly Gly Arg Arg Cys Gly
                410                 415                 420
Asn Gly Tyr Leu Glu Asp Gly Glu Glu Cys Asp Cys Gly Glu Glu
                425                 430                 435
Glu Glu Cys Asn Asn Pro Cys Cys Asn Ala Ser Asn Cys Thr Leu
                440                 445                 450
Arg Pro Gly Ala Glu Cys Ala His Gly Ser Cys Cys His Gln Cys
                455                 460                 465
Lys Leu Leu Ala Pro Gly Thr Leu Cys Arg Glu Gln Ala Arg Gln
                470                 475                 480
Cys Asp Leu Pro Glu Phe Cys Thr Gly Lys Ser Pro His Cys Pro
                485                 490                 495
Thr Asn Phe Tyr Gln Met Asp Gly Thr Pro Cys Glu Gly Gly Gln
                500                 505                 510
Ala Tyr Cys Tyr Asn Gly Met Cys Leu Thr Tyr Gln Glu Gln Cys
                515                 520                 525
Gln Gln Leu Trp Gly Pro Gly Ala Arg Pro Ala Pro Asp Leu Cys
                530                 535                 540
Phe Glu Lys Val Asn Val Ala Gly Asp Thr Phe Gly Asn Cys Gly
                545                 550                 555
Lys Asp Met Asn Gly Glu His Arg Lys Cys Asn Met Arg Asp Ala
                560                 565                 570
Lys Cys Gly Lys Ile Gln Cys Gln Ser Ser Glu Ala Arg Pro Leu
                575                 580                 585
Glu Ser Asn Ala Val Pro Ile Asp Thr Thr Ile Ile Met Asn Gly
                590                 595                 600
Arg Gln Ile Gln Cys Arg Gly Thr His Val Tyr Arg Gly Pro Glu
                605                 610                 615
Glu Glu Gly Asp Met Leu Asp Pro Gly Leu Val Met Thr Gly Thr
                620                 625                 630
Lys Cys Gly Tyr Asn His Ile Cys Phe Glu Gly Gln Cys Arg Asn
                635                 640                 645
Thr Ser Phe Phe Glu Thr Glu Gly Cys Gly Lys Lys Cys Asn Gly
                650                 655                 660
His Gly Val Cys Asn Asn Asn Gln Asn Cys His Cys Leu Pro Gly
                665                 670                 675
Trp Ala Pro Pro Phe Cys Asn Thr Pro Gly His Gly Gly Ser Ile
                680                 685                 690
```

```
Asp Ser Gly Pro Met Pro Pro Glu Ser Val Gly Pro Val Val Ala
                695                 700                 705
Gly Val Leu Val Ala Ile Leu Val Leu Ala Val Leu Met Leu Met
                710                 715                 720
Tyr Tyr Cys Cys Arg Gln Asn Asn Lys Leu Gly Gln Leu Lys Pro
                725                 730                 735
Ser Ala Leu Pro Ser Lys Leu Arg Gln Gln Phe Ser Cys Pro Phe
                740                 745                 750
Arg Val Ser Gln Asn Ser Gly Thr Gly His Ala Asn Pro Thr Phe
                755                 760                 765
Lys Leu Gln Thr Pro Gln Gly Lys Arg Lys Val Ile Asn Thr Pro
                770                 775                 780
Glu Ile Leu Arg Lys Pro Ser Gln Pro Pro Pro Arg Pro Pro Pro
                785                 790                 795
Asp Tyr Leu Arg Gly Gly Ser Pro Pro Ala Pro Leu Pro Ala His
                800                 805                 810
Leu Ser Arg Ala Ala Arg Asn Ser Pro Gly Pro Gly Ser Gln Ile
                815                 820                 825
Glu Arg Thr Glu Ser Ser Arg Arg Pro Pro Pro Ser Arg Pro Ile
                830                 835                 840
Pro Pro Ala Pro Asn Cys Ile Val Ser Gln Asp Phe Ser Arg Pro
                845                 850                 855
Arg Pro Pro Gln Lys Ala Leu Pro Ala Asn Pro Val Pro Gly Arg
                860                 865                 870
Arg Ser Leu Pro Arg Pro Gly Gly Ala Ser Pro Leu Arg Pro Pro
                875                 880                 885
Gly Ala Gly Pro Gln Gln Ser Arg Pro Leu Ala Ala Leu Ala Pro
                890                 895                 900
Lys Phe Pro Glu Tyr Arg Ser Gln Arg Ala Gly Gly Met Ile Ser
                905                 910                 915
Ser Lys Ile
```

<210> 19
<211> 218
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4847254CD1

<400> 19

```
Met Arg Gln Gly Pro Tyr Leu Pro Leu Glu Leu Gly Leu Glu Gln
  1               5                  10                  15
Leu Phe Gln Glu Leu Ala Gly Glu Glu Glu Glu Leu Asn Ala Ser
                 20                  25                  30
Gln Leu Gln Ala Leu Leu Ser Ile Ala Leu Glu Pro Ala Arg Ala
                 35                  40                  45
His Thr Ser Thr Pro Arg Glu Ile Gly Leu Arg Thr Cys Glu Gln
                 50                  55                  60
Leu Leu Gln Cys Phe Gly Val His Gly Gly Gln Cys Leu Gly Glu
                 65                  70                  75
Gly Gly Ser Gly Glu Gly Asp Val Gly Val Ser Pro Pro Leu Leu
                 80                  85                  90
Glu Arg Leu Thr Leu Thr Arg Cys Pro Arg Pro Pro Thr Gln His
```

```
                95                  100                 105
Gly Gln Ser Leu Ala Leu His His Phe Gln Gln Leu Trp Gly Tyr
                110                 115                 120
Leu Leu Glu Trp Gln Ala Ile Phe Asn Lys Phe Asp Glu Asp Thr
                125                 130                 135
Ser Gly Thr Met Asn Ser Tyr Glu Leu Arg Leu Ala Leu Asn Ala
                140                 145                 150
Ala Gly Phe His Leu Asn Asn Gln Leu Thr Gln Thr Leu Thr Ser
                155                 160                 165
Arg Tyr Arg Asp Ser Arg Leu Arg Val Asp Phe Glu Arg Phe Val
                170                 175                 180
Ser Cys Val Ala His Leu Thr Cys Ile Phe Cys His Cys Ser Gln
                185                 190                 195
His Leu Asp Gly Gly Glu Gly Val Ile Cys Leu Thr His Arg Gln
                200                 205                 210
Trp Met Glu Val Ala Thr Phe Ser
                215
```

<210> 20
<211> 656
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5776350CD1

<400> 20

```
Met Lys Leu Glu Pro Leu Gln Glu Arg Glu Pro Ala Pro Glu Glu
  1                 5                  10                  15
Asn Leu Thr Trp Ser Ser Ser Gly Gly Asp Glu Lys Val Leu Pro
                  20                  25                  30
Ser Ile Pro Leu Arg Cys His Ser Ser Ser Ser Pro Val Cys Pro
                  35                  40                  45
Arg Arg Lys Pro Arg Pro Arg Pro Gln Pro Arg Ala Arg Ser Arg
                  50                  55                  60
Ser Gln Pro Gly Leu Ser Ala Pro Pro Pro Pro Pro Ala Arg Pro
                  65                  70                  75
Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Ala Pro Arg Pro Arg
                  80                  85                  90
Ala Trp Arg Gly Ser Arg Arg Arg Ser Arg Pro Gly Ser Arg Pro
                  95                 100                 105
Gln Thr Arg Arg Ser Cys Ser Gly Asp Leu Asp Gly Ser Gly Asp
                 110                 115                 120
Pro Gly Gly Leu Gly Asp Trp Leu Leu Glu Val Glu Phe Gly Gln
                 125                 130                 135
Gly Pro Thr Gly Cys Ser His Val Glu Ser Phe Lys Val Gly Lys
                 140                 145                 150
Asn Trp Gln Lys Asn Leu Arg Leu Ile Tyr Gln Arg Phe Val Trp
                 155                 160                 165
Ser Gly Thr Pro Glu Thr Arg Lys Arg Lys Ala Lys Ser Cys Ile
                 170                 175                 180
Cys His Val Cys Ser Thr His Met Asn Arg Leu His Ser Cys Leu
                 185                 190                 195
Ser Cys Val Phe Phe Gly Cys Phe Thr Glu Lys His Ile His Lys
                 200                 205                 210
```

```
His Ala Glu Thr Lys Gln His His Leu Ala Val Asp Leu Tyr His
                215                 220                 225
Gly Val Ile Tyr Cys Phe Met Cys Lys Asp Tyr Val Tyr Asp Lys
                230                 235                 240
Asp Ile Glu Gln Ile Ala Lys Glu Thr Lys Glu Lys Ile Leu Arg
                245                 250                 255
Leu Leu Thr Ser Thr Ser Thr Asp Val Ser His Gln Gln Phe Met
                260                 265                 270
Thr Ser Gly Phe Glu Asp Lys Gln Ser Thr Cys Glu Thr Lys Glu
                275                 280                 285
Gln Glu Pro Lys Leu Val Lys Pro Lys Lys Lys Arg Arg Lys Lys
                290                 295                 300
Ser Val Tyr Thr Val Gly Leu Arg Gly Leu Ile Asn Leu Gly Asn
                305                 310                 315
Thr Cys Phe Met Asn Cys Ile Val Gln Ala Leu Thr His Ile Pro
                320                 325                 330
Leu Leu Lys Asp Phe Phe Leu Ser Asp Lys His Lys Cys Ile Met
                335                 340                 345
Thr Ser Pro Ser Leu Cys Leu Val Cys Glu Met Ser Ser Leu Phe
                350                 355                 360
His Ala Met Tyr Ser Gly Ser Arg Thr Pro His Ile Pro Tyr Lys
                365                 370                 375
Leu Leu His Leu Ile Trp Ile His Ala Glu His Leu Ala Gly Tyr
                380                 385                 390
Arg Gln Gln Asp Ala His Glu Phe Leu Ile Ala Ile Leu Asp Val
                395                 400                 405
Leu His Arg His Ser Lys Asp Asp Ser Gly Gly Gln Glu Ala Asn
                410                 415                 420
Asn Pro Asn Cys Cys Asn Cys Ile Ile Asp Gln Ile Phe Thr Gly
                425                 430                 435
Gly Leu Gln Ser Asp Val Thr Cys Gln Ala Cys His Ser Val Ser
                440                 445                 450
Thr Thr Ile Asp Pro Cys Trp Asp Ile Ser Leu Asp Leu Pro Gly
                455                 460                 465
Ser Cys Ala Thr Phe Asp Ser Gln Asn Pro Glu Arg Ala Asp Ser
                470                 475                 480
Thr Val Ser Arg Asp Asp His Ile Pro Gly Ile Pro Ser Leu Thr
                485                 490                 495
Asp Cys Leu Gln Trp Phe Thr Arg Pro Glu His Leu Gly Ser Ser
                500                 505                 510
Ala Lys Ile Lys Cys Asn Ser Cys Gln Ser Tyr Gln Glu Ser Thr
                515                 520                 525
Lys Gln Leu Thr Met Lys Lys Leu Pro Ile Val Ala Cys Phe His
                530                 535                 540
Leu Lys Arg Phe Glu His Val Gly Lys Gln Arg Arg Lys Ile Asn
                545                 550                 555
Thr Phe Ile Ser Phe Pro Leu Glu Leu Asp Met Thr Pro Phe Leu
                560                 565                 570
Ala Ser Thr Lys Glu Ser Arg Met Lys Glu Gly Gln Pro Pro Thr
                575                 580                 585
Asp Cys Val Pro Asn Glu Asn Lys Tyr Ser Leu Phe Ala Val Ile
                590                 595                 600
Asn His His Gly Thr Leu Glu Ser Gly His Tyr Thr Ser Phe Ile
                605                 610                 615
Arg Gln Gln Lys Asp Gln Trp Phe Ser Cys Asp Asp Ala Ile Ile
                620                 625                 630
```

```
Thr Lys Ala Thr Ile Glu Asp Leu Leu Tyr Ser Glu Gly Tyr Leu
                635                 640                 645
Leu Phe Tyr His Lys Gln Gly Leu Glu Lys Asp
                650                 655
```

<210> 21
<211> 509
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473300CD1

<400> 21

```
Met Leu Leu Thr Gln Ser Leu Phe Gly Gly Leu Phe Thr Arg Thr
  1               5                  10                  15
Arg Glu Thr Val Cys Ile Phe Gln Pro Trp Thr Gln Gln Arg Val
             20                  25                  30
Thr Thr Asn Arg Ser Trp Thr His Pro Glu Thr Gln Ala Glu Arg
             35                  40                  45
Leu Trp Ile Lys Gln Glu Thr Glu Asp Arg Asp Arg Ser Ser Phe
             50                  55                  60
Tyr Ile Gln Met Asn Lys Gly Arg Pro Trp Val Tyr Leu Lys Tyr
             65                  70                  75
Gln Ile Val Gly Ala Trp Ile Gln Pro Glu Leu Asp Val Ile His
             80                  85                  90
Ser Phe Ile Gln Ser Glu Thr Phe Leu Leu Arg Phe Trp Pro Lys
             95                 100                 105
Val Leu Ser Pro Val Val Lys Pro Trp Ile Leu Leu Lys Gly Arg
            110                 115                 120
Thr Leu Ile Ser Trp Ile Leu Pro Val Thr Arg Ala Asp Thr Gly
            125                 130                 135
Ser Ser Leu Lys Phe Ile Leu Leu Asn Pro Ser Val Phe Leu Lys
            140                 145                 150
Pro Ala Asn His Leu Ser Thr Trp Asp Arg Arg His Thr Leu Leu
            155                 160                 165
His Leu Asp Asn Phe Val Val Val Val Leu Ala Val Glu Ser Pro
            170                 175                 180
Gly Ile Val Gln Lys Arg His Leu Ser Ile Leu Gln Val Ser Thr
            185                 190                 195
Cys Ala Gln Phe Trp Leu Lys Leu Asn Glu Leu Thr Phe Trp Val
            200                 205                 210
Glu Ala Lys Lys Ala Met Trp Met Ala Asp Tyr Gln Gly Val Thr
            215                 220                 225
Gln Ser Ser Tyr Ala Pro Trp Tyr Lys Gln Gly Pro Met Thr Thr
            230                 235                 240
Ser Ala Ser Met Ser His Ser Val Ser Thr Ser Thr Asn Ala Ser
            245                 250                 255
Ala Phe Thr Ser Thr Pro Ala Ser Leu Trp Pro His Phe Ser Leu
            260                 265                 270
Pro Gln Pro Gln Ser Lys Ala Gln Lys Leu Gly Arg Asp Gln Ile
            275                 280                 285
Tyr Leu Arg Tyr Ala Met Pro Trp Lys Ala Val Ile Ile Ile Cys
            290                 295                 300
Gly Ser Gln Ile Cys Ser Gly Ser Ile Val Gly Ser Ser Trp Ile
```

```
                305                 310                 315
Leu Thr Ala Ala His Cys Val Arg Lys Leu Arg Asp Pro Glu Asp
                320                 325                 330
Thr Ala Val Ile Leu Gly Leu Arg His Pro Gly Ala Pro Leu Arg
                335                 340                 345
Val Val Lys Val Ser Thr Ile Leu Leu His Glu Arg Phe Trp Leu
                350                 355                 360
Val Thr Glu Ala Ala Arg Asn Ile Leu Glu Leu Leu Leu Leu His
                365                 370                 375
Asp Val Gln Thr Pro Ile Trp Leu Leu Ser Leu Leu Gly Tyr Leu
                380                 385                 390
Arg Asn Leu Asn Ser Ser Glu Cys Trp Leu Ser Arg Pro His Ile
                395                 400                 405
Val Thr Pro Ala Val Leu Leu Arg His Pro Trp Ala Pro Gly Gly
                410                 415                 420
Pro Gln Pro His Pro Gly Thr Gly Pro Leu Pro Gln Ile Gln Ala
                425                 430                 435
Gln Gln Pro Asn Leu Gln Ile His His Val Ala Gln Gln Asp Phe
                440                 445                 450
Ile Ile Cys Asp Pro Gly Pro Tyr Leu Gly Pro Ser Leu Glu His
                455                 460                 465
His Val Phe Leu Gly Trp Leu Pro Ala Thr Leu Leu Leu Gly Pro
                470                 475                 480
Arg Arg Pro Pro Pro Ala Ala Ser His Pro Glu Leu Ala Ala Ala
                485                 490                 495
Lys Thr Trp Leu Trp Pro Gly Asn Arg Gly Cys Pro Val Ala
                500                 505
```

<210> 22
<211> 2789
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5155802CB1

<400> 22
ctctttctct ctccctctgg catgcatgct gctggtagga gacccccaag tcaacattgc 60
ttcagaaatc ctttagcact catttctcag gagaacttat ggcttcagaa tcacagctcg 120
gtttttaaga tggacataac ctgtacgacc ttctgatggg ctttcaactt tgaactggat 180
gtggacactt ttctctcaga tgacagaatt actccaactt ccccttttgca gttgcttcct 240
ttccttgaag gtagctgtat cttatttttct ttaaaaagct ttttcttcca aagccacttg 300
ccatgccgac cgtcattagc gcatctgtgg ctccaaggac agcggctgag ccccggtccc 360
cagggccagt tcctcacccg gcccagagca aggccactga ggctgggggt ggaaacccaa 420
gtggcatcta ttcagccatc atcagccgca attttcctat tatcggagtg aaagagaaga 480
cattcgagca acttcacaag aaatgtctag aaaagaaagt tctttatgtg gaccctgagt 540
tcccaccgga tgagacctct ctcttttata gccagaagtt ccccatccag ttcgtctgga 600
agagacctcc ggaaatttgc gagaatcccc gatttatcat tgatggagcc aacagaactg 660
acatctgtca aggagagcta ggggactgct ggtttctcgc agccattgcc tgcctgaccc 720
tgaaccagca ccttcttttc cgagtcatac cccatgatca aagtttcatc gaaaactacg 780
cagggatctt ccacttccag ttctggcgct atggagagtg ggtggacgtg gttatagatg 840
actgcctgcc aacgtacaac aatcaactgg ttttcaccaa gtccaaccac cgcaatgagt 900
tctggagtgc tctgctggag aaggcttatg ctaagctcca tggttcctac gaagctctga 960
aaggtgggaa caccacagag gccatggagg acttcacagg aggggtgaca gagtttttg 1020
agatcaggga tgctcctagt gacatgtaca agatcatgaa gaaagccatc gagagaggct 1080

ccctcatggg ctgctccatt gatacaatca ttccggttca gtatgagaca agaatggcct 1140
gcgggctggt cagaggtcac gcctactctg tcacggggct ggatgaggtc ccgttcaaag 1200
gtgagaaagt gaagctggtg cggctgcgga atccgtgggg ccaggtggag tggaacggtt 1260
cttggagtga tagatggaag gactggagct ttgtggacaa agatgagaag gcccgtctgc 1320
agcaccaggt cactgaggat ggagagttct ggatgtccta tgaggatttc atctaccatt 1380
tcacaaagtt ggagatctgc aacctcacgg ccgatgctct gcagtctgac aagcttcaga 1440
cctggacagt gtctgtgaac gagggccgct gggtacgggg ttgctctgcc ggaggctgcc 1500
gcaacttccc agatactttc tggaccaacc ctcagtaccg tctgaagctc ctggaggagg 1560
acgatgaccc tgatgactcg gaggtgattt gcagcttcct ggtggccctg atgcagaaga 1620
accggcggaa ggaccggaag ctaggggcca gtctcttcac cattggcttc gccatctacg 1680
aggttcccaa agagatgcac gggaacaagc agcacctgca gaaggacttc ttcctgtaca 1740
acgcctccaa ggccaggagc aaaacctaca tcaacatgcg ggaggtgtcc cagcgcttcc 1800
gcctgcctcc cagcgagtac gtcatcgtgc cctccaccta cgagccccac caggaggggg 1860
aattcatcct ccgggtcttc tctgaaaaga ggaacctctc tgaggaagtt gaaaatacca 1920
tctccgtgga tcggccagtg cccatcatct tcgtttcgga cagagcaaac agcaacaagg 1980
agctgggtgt ggaccaggag tcagaggagg gcaaaggcaa aacaagccct gataagcaaa 2040
agcagtcccc acagccacag cctggcagct ctgatcagga aagtgaggaa cagcaacaat 2100
tccggaacat tttcaagcag atagcaggag atgacatgga gatctgtgca gatgagctca 2160
agaaggtcct taacacagtc gtgaacaaac acaaggacct gaagacacac gggttcacac 2220
tggagtcctg ccgtagcatg attgcgctca tggatacaga tggctctgga aagctcaacc 2280
tgcaggagtt ccaccacctc tggaacaaga ttaaggcctg gcagaaaatt ttcaaacact 2340
atgacacaga ccagtccggc accatcaaca gctacgagat gcgaaatgca gtcaacgacg 2400
caggattcca cctcaacaac cagctctatg acatcattac catgcggtac gcagacaaac 2460
acatgaacat cgactttgac agtttcatct gctgcttcgt taggctggag ggcatgttca 2520
gagcttttca tgcatttgac aaggatggag atggtatcat caagctcaac gttctggagt 2580
ggctgcagct caccatgtat gcctgaacca ggctggcctc atccaaagcc atgcaggatc 2640
actcaggatt tcagtttcac cctctatttc caaagccatt tacctcaaag gacccagcag 2700
ctacacccct acaggcttcc aggcacctca tcagtcatgt tcctcctcca ttttaccccc 2760
tacccatcct tgatcggtca tgcctagcc 2789

<210> 23
<211> 2267
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71269782CB1

<400> 23
gtaagtgaca caacttgaaa ctgcttggcc ctctttaaaa agaaataata aaatgggaga 60
gaatgaagca agtttaccta acacgtcttt gcaaggtaaa aagatggcct atcagaaggt 120
ccatgcagat caaagagctc caggacactc acagtactta gacaatgatg accttcaagc 180
cactgcccct gacttagagt gggacatgga gaaggaacta gaggagtctg gttttgacca 240
attccagcta gacggtgctg agaatcagaa cctagggcat tcagagacta tagacctcaa 300
tcttgattcc attcaaccag caacttcacc caaaggaagg ttccagagac ttcaagaaga 360
atctgactac attacccatt atacacgatc tgcaccaaag agcaatcgct gcaacttttg 420
ccacgtctta aaaatacttt gcacagccac cattttattt atttttggga ttttgatagg 480
ttattatgta catacaaatt gcccttcaga tgctccatct tcaggaacag ttgatcctca 540
gttatatcaa gagattctca agacaatcca ggcagaagat attaagaagt ctttcagaaa 600
tttggtacaa ctatataaaa atgaagatga catggaaatt tcaaagaaga ttaagactca 660
gtggacctct ttgggcctag aagatgtaca gtttgtaaat tactctgtgc tgcttgatct 720
gccaggccct tctcccagca ctgtgactct gagcagcagt ggtcaatgct ttcatcctaa 780
tggccagcct tgcagtgaag aagccagaaa agatagcagc caagacctgc tctattcata 840
tgcagcctat tctgccaaag gaactctcaa ggctgaagtc atcgatgtga gttatggaat 900
ggcagatgat ttaaaaagga ttaggaaaat aaaaaacgta acaaatcaga tcgcactcct 960

gaaattagga aaattgccac tgctttataa gctttcctca ttggaaaagg ctggatttgg 1020
aggtgttctt ctgtatatcg atccttgtga tttgccaaag actgtgaatc ctagccatga 1080
taccttcatg gtgtcactga atccaggagg agacccttct acgcctggtt acccaagtgt 1140
cgatgaaagt tttagacaaa gccgatcaaa cctcacctct ctattagtgc agcccatctc 1200
tgcatccctc gttgcaaaac tgatctcttc gccaaaagct agaaccaaaa atgaagcgtg 1260
tagctctcta gagcttccaa ataatgaaat aagagtcgtc agcatgcaag ttcagacagt 1320
cacaaaattg aaaacagtta ctaatgttgt tggatttgta atgggcttga catctccaga 1380
ccggtatatc atagttggca gccatcatca cactgcacac agttataatg gacaagaatg 1440
ggccagtagt actgcaataa tcacagcgtt tatccgtgcc ttgatgtcaa aagttaagag 1500
agggtggaga ccagaccgaa ctattgtttt ctgttcttgg ggaggaacag cttttggcaa 1560
tattggctca tatgaatggg gagaggattt caagaaggtt cttcaaaaaa atgttgtggc 1620
ttatattagc ctccacagtc ccataagggg gaactctagt ctgtatcctg tagcatcacc 1680
atctcttcag caactggtag tagaggtaag acaaaccact attgtatcaa atgattatgc 1740
aaaaccgacc ttttctctat attttgacat ttcttgattt tttcatttat ttttaaatat 1800
gcatcaaatg ttgtataagt gttttaagaa atgatctatt gctgacattt tatcaatata 1860
ccttaactaa tttcttgtgt tctggaattc ttcacttgct actcttttat ggtcatattt 1920
ctagaagaca tgagtcacac agttatagag aaggtataca aaaatatatt tttaaaaaat 1980
atatgaattt agctctcaaa ttcccaattc tgtaatcttg acattttatg ataagcctgg 2040
ttacttttga atttcttcct cttcattctt gttttaagta aatgtgagac ctgtcctatc 2100
tttacaactg ctgtgtaggc cccccgagag caagaatata gtgataacta aatttaaaag 2160
atttagaaaa tattgtttga aaaattacct gtggaaaaag aaaacatgtt ttcttagtat 2220
cctgaaaaat catatatttt ttatgtttca ttggagttac ttatttt 2267

<210> 24
<211> 963
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472651CB1

<400> 24
atgggggacc cagaaggaag cgcagagtgg ggttggggga aggggatacc ggtggtcaga 60
agaaatttat taacagtgga tgggataagt ctgtgtctgg agggatcctg gtggaggcag 120
aagggtcctg cctcacctgg attctctcac tccctcccca gactgcagcc gaaccctggt 180
ccctcctcca caatgtggct tctcctcact ctctccttcc tgctggcatc cacagcagcc 240
caggatggtg acaagttgct ggaaggtgac gagtgtgcac cccactccca gccatggcaa 300
gtggctctct acgagcgtgg acgctttaac tgtggcgctt ccctcatctc cccacactgg 360
gtgctgtctg cggcccactg ccaaagccgc ttcatgagag tgcgcctggg agagcacaac 420
ctgcgcaagc gcgatggccc agagcaacta cggaccacgt ctcgggtcat tccacacccg 480
cgctacgaag cgcgcagcca ccgcaacgac atcatgttgc tgcgcctagt ccagcccgca 540
cgcctgaacc cccaggtgcg ccccgcggtg ctacccacgc gttgccccca cccgggggag 600
gcctgtgtgg tgtctggctg gggcctggtg tcccacaacg agcctgggac cgctgggagc 660
ccccggtcac aagtgagtct cccagatacg ttgcattgtg ccaacatcag cattatctcg 720
gacacatctt gtgacaagag ctacccaggg cgcctgacaa acaccatggt gtgtgcaggc 780
gcggagggca gaggcgcaga atcctgtgag ggtgactctg ggggacccct ggtctgtggg 840
ggcatcctgc agggcattgt gtcctggggt gacgtccctt gtgacaacac caccaagcct 900
ggtgtctata ccaaagtctg ccactacttg gagtggatca gggaaaccat gaagaggaac 960
tga 963

<210> 25
<211> 1137
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7478251CB1

<400> 25
atggctgaga aaccatccaa cggtgttctg gtccacatgg tgaagttgct gatcaagacc 60
tttctagatg gcatttttga tgatttgatg gaaaataatg tattaaatac agatgagata 120
caccttatag gaaaatgtct aaagtttgtg gtgagcaatg ctgaaaacct ggttgatgat 180
atcactgaga cagctcaaac tgcaggcaaa atatttaggg aacacctgtg gaattccaaa 240
aaacagctga gttcaatttt tttctctctt tcagcttttc tggaaatcca gggtgcccaa 300
cccagtggca agttaaagct ttgtcctcat gctcacttcc atgaactaaa gacaaaaagg 360
gcagatgaga tatatccagt gatggagaaa aaaaggcgaa catgcctggg cctcaacatc 420
cgcaacaaag aattcaacta tcttcataat cgaaatggtt ctgaacttga ccttttgggg 480
atgcgagatc tacttgaaaa ccttggatac tcagtggtta taaaagagaa tctcacagct 540
caggaaatgg aaacagcact aaggcagttt gctgctcacc cagagcacca gtcctcagac 600
agcacattcc tggtgtttat gtcacatagc atcctgaatg gaatctgtgg gaccaagcac 660
tgggatcaag agccagatgt tcttcacgat gacaccatct ttgaaatttt caacaaccgt 720
aactgccaga gtctgaaaga caaacccaag gtcatcatca tgcaagcctg ccgaggcaat 780
ggtgctggga ttgtttggtt caccactgac agtggaaaag ccggtgcaga tactcatggt 840
cggctcttgc aaggtaacat ctgtaatgat gctgttacaa aggctcatgt ggaaaaggac 900
ttcattgctt tcaaatcttc cacaccacat aatgtttctt ggagacatga aacaaatggc 960
tctgtcttca tttcccaaat tatctactac ttcagagagt attcttggag tcatcatcta 1020
gaggaaattt ttcaaaaggt acaacattca tttgagaccc caaatatact gacccagctg 1080
cccaccattg aaagactatc catgacacga tatttctatc tctttcctgg gaattaa 1137

<210> 26
<211> 3204
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2759385CB1

<400> 26
gccagcgcgc caccatgggc agtcccggtt tccccttgta aagatggcgg tgagggatcg 60
ctgcaacctt tagactaatg actgtccgaa acatcgcctc catctgtaat atgggcacca 120
atgcctctgc tctggaaaaa gacattggtc cagagcagtt tccaatcaat gaacactatt 180
tcggattggt caattttgga aacacatgct actgtaactc cgtgcttcag gcattgtact 240
tctgccgtcc attccgggag aatgtgttgg catacaaggc ccagcaaaag aagaaggaaa 300
acttgctgac gtgcctggcg gacctttttcc acagcattgc cacacagaag aagaaggttg 360
gcgtcatccc accaaagaag ttcatttcaa ggctgagaaa agagaatgat ctctttgata 420
actacatgca gcaggatgct catgaatttt taaattattt gctaaacact attgcggaca 480
tccttcagga ggagaagaaa caggaaaaac aaaatggaaa attaaaaaat ggcaacatga 540
acgaacctgc ggaaaataat aaaccagaac tcacctgggt ccatgagatt tttcagggaa 600
cgcttaccaa tgaaactcga tgcttgaact gtgaaactgt tagtagcaaa gatgaagatt 660
ttcttgacct ttctgttgat gtggagcaga atacatccat tacccactgt ctaagagact 720
tcagcaacac agaaacactg tgtagtgaac aaaaatatta ttgtgaaaca tgctgcagca 780
aacaagaagc ccagaaaagg atgagggtaa aaaagctgcc catgatcttg gccctgcacc 840
taaagcggtt caagtacatg gagcagctgc acagatacac caagctgtct taccgtgtgg 900
tcttccctct ggaactccgg ctcttcaaca cctccagtga tgcagtgaac ctggaccgca 960
tgtatgactt ggttgcggtg gtcgttcact gtggcagtgg tcctaatcgt gggcattata 1020
tcactattgt gaaaagtcac ggcttctggc ttttgtttga tgatgacatt gtagagaaaa 1080
tagatgctca agctattgaa gaattctatg gcctgacgtc agatatatca aaaaattcag 1140
aatctggata tattttattc tatcagtcaa gagagtaact gaaagacctg cgggactgat 1200
tcacgtgggg agaatgttca cagcactgtc acccggcttc tccgcaggct ttcctcttcc 1260

ccagtggccc actaatggta tcactccgag tctcaatggt ctggctgtgt tagactctct 1320
ccttttgtgt ttttacatgc agcactactc ttggttttat ttcagtctga catagagtta 1380
actgcaatca gattgtagtc tgatttatat gaataacggt tgctaatttt aggactgggt 1440
gaaagctatg ccattcatta tgtctggctg tattagaatg acatttccta tgaatgtcta 1500
cggtctgttt taggtgtttg ctaaacttct atggcttcca gggtcttctt acaatgcatt 1560
cctttaactt gtccctggaa gcattgctac ccattttcag cttctctgcc tctcttctga 1620
tacaaggaca gaagaattgg gtagatattc acctttttagg ggtgcaagta tagctttaag 1680
tttgtgcaag tgaaaatgtt gaaaagtgag taacctcgat attaaaatca tccttgacat 1740
gaaacagggt gaagagaagc tgtccgtggc ggctggtgtt ggctggcatt ggcactgggc 1800
tgtgctgacc tagccattac aattccaggg gctaagaagg ctcagggcag acaaagtcaa 1860
gaggaggaag tttttgtgga caatgaaaag ttatttttcgt acctttctac caaaaccaag 1920
tttcaggaaa ataactctat gttgtttatt ttcagtgaca cttatgtaag gccctgtgag 1980
ttgtatttat cctgtatccg gcactgctaa gcttttcaag gtatctttcc aatcctgctg 2040
atgtggcagt caatggctgc agggctggca aacctcccct tagccagtca gcacggcatt 2100
gttccttatc aggaataaca aaggtactac atctttccag ccatcagcac gttgtacaac 2160
ttaacttttt aacatagtcc gtctgtttac tgaggcactg gcgagtccca gggctgatac 2220
agaaccttcc ctagagggaa taccagagtt agctgggtat agaggtggct caaaggaagt 2280
gtccgtgggc agggggagga atgaacaaaa tggcgctgtt tctttggctc agactcctag 2340
aatgcttgac aagacagaat ttttttggaa gaacctcatc tcactatagt tactttttc 2400
acttttgtta tatatgtatt tattagagca tttgaatatt ggtaccttta aaagggtcat 2460
ttggtgtttt gctgttgagc tggtttttga gtcatagatc ttggcttcct ttagaagcca 2520
cttaacttcc atacactata ataaactgtg aacatatttt tgttacctaa tgcatccact 2580
gatgaacatg caaactttgg gcataatgtg aactaaaatt gaaatggaaa atgttagtgg 2640
ccattttgca acaatgaaga ggatagcact ttatctagat gaaaactgga tttcttatct 2700
ttgaaatatc ttgaactgtt tattgctcag aacttaagta agcatgccaa catttcgttt 2760
gtttatgctt gaagtgaaat gttttacttt tcactggaga agacaaaaca gggtgatctt 2820
catgttattg ttttatacaa gtgatggaaa tgtaccttgc cttgtttaga ggcaatttca 2880
catttataaa tatttttttt tcctccatga aacttacgca gtaatcacta cctggaaggt 2940
gagttttgat ctcttttttaa ggagaggcac tttccaactg aaggtgattg atgtagggaa 3000
atgtttgtac tatatagaat ccatatattt gactgcaagt tacaaagttt taagaacatg 3060
atggttggtc tctaatatat ttggaactga ttcataagaa aagttattaa aattatcttt 3120
gaaacacctc ttgaagctaa tttattagaa aaaatatttc agttggaagg ctgtagaagt 3180
aatgtttaaa tgctaagtca taag 3204

<210> 27
<211> 1641
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4226182CB1

<400> 27
attttaatct atacattgaa acgatttgtc actgtcactc aacaaagtat tttttatcag 60
aatattggag caaagccttt ggcaaacata gccagatgtg gtgagaacac taaaggcatt 120
aaaaactttg atctattaga tatgtttcag atatcaagag tgtttaatct aattaatact 180
aatatgtcat attaaataat attccaagtt tgaaacaatt gaggacatat ggaaagatca 240
tacctcaatt tgcttcagat ttggatttta tgaactgcag acttaaatta ttagcaggaa 300
ttctcatttt taaattgtct gttaaaatca attataaatg taaatttatt tatttagtta 360
tatggattat cctcgttatt tgggagcagt gtttcctgga acaatgtgta ttactcgtta 420
ttctgcagga gttgcattgc aatgtggacc tgcaagctgt tgtgattttc gaacttgtgt 480
actgaaagac ggagcaaaat gttataaagg actgtgctgc aaagactgtc aaatttaca 540
atcaggcgtt gaatgtaggc cgaaagcaca tcctgaatgt gacatcgctg aaaattgtaa 600
tggaagctca ccagaatgtg gtcctgacat aactttaatc aatggacttt catgcaaaaa 660
taataagttt atttgttatg acggagactg ccatgatctc gatgcacgtt gtgagagtgt 720

atttggaaaa ggttcaagaa atgctccatt tgcctgctat gaagaaatac aatctcaatc 780
agacagattt gggaactgtg gtagggatag aaataacaaa tatgtgttct gtggatggag 840
gaatcttata tgtggaagat tagtttgtac ctaccctact cgaaagcctt tccatcaaga 900
aaatggtgat gtgatttatg ctttcgtacg agattctgta tgcataactg tagactacaa 960
attgcctcga acagttccag atccactggc tgtcaaaaat ggctctcagt gtgatattgg 1020
gagggtttgt gtaaatcgtg aatgtgtaga atcaaggata attaaggctt cagcacatgt 1080
ttgttcacaa cagtgttctg gacatggagt gtgtgattcc agaaacaagt gccattgttc 1140
gccaggctat aagcctccaa actgccaaat acgttccaaa ggattttcca tatttcctga 1200
ggaagatatg ggttcaatca tggaaagagc atctgggaag actgaaaaca cctggcttct 1260
aggtttcctc attgctcttc ctattctcat tgtaacaacc gcaatagttt tggcaaggaa 1320
acagttgaaa aagtggttcg ccaaggaaga ggaattccca agtagcgaat ctaaatcgga 1380
aggtagcaca cagacatatg ccagccaatc cagctcagaa ggcagcactc agacatatgc 1440
cagccaaacc agatcagaaa gcagcagtca agctgatact agcaaatcca aatcacagga 1500
cagtacccaa acacaaagca gtagtaacta gtgattcctt cagaaggcaa cggataacat 1560
cgagagtctc gctaagaaat gaaaattctg tctttccttc cgtggtcaca gctgaaagaa 1620
acaataaatt gagtgtggat c 1641

<210> 28
<211> 1983
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5078962CB1

<400> 28
cctgattggt gctaccaagc ccaaaataga cccatggtaa aacaccaaat tcggccatgc 60
agcaacccct gtaatgtttt actgagataa aaaagatgcg agttgacctą cttttcaagc 120
ttttcaaggt tcaactgatg aacctttga gcatttattc acatgcgctg ggtagcccag 180
ggcaccaatc attgagaaag agtaaggaat tgccgaagaa cataattttg aaatcctcag 240
gccaaaaggg agttatgtca tttaatgact cacaaatgat ttagaggatc gtagggttta 300
acatttctat ttcctaatgg tccataacac catcatatgc ccaaatgatt gtccacaagg 360
cacagttgag gattctaaca ctaatcataa ttaattcaaa tgttgtacca taactttatc 420
atagtaaatt tatacagtct cacatggaag tactgttgct atagcatagt tgataaatac 480
aagaaatgtc ttcaattgtt gctgcacaat ttctttattt aacattttag gttgacatga 540
caactgaaga gatagatgct cttgttcatc gggaaatcat cagtcataat gcctatccct 600
cacctctagg ctatggaggt tttccaaaat ctgtttgtac ctctgtaaac aacgtgctct 660
gtcatggtat tcctgacagt cgacctcttc aggatggaga tattatcaac attgatgtca 720
cagtctatta caatggctac catggagaca cctctgaaac atttttggtg ggcaatgtgg 780
acgaatgtgg taaaaagtta gtggaggttg ccaggaggtg tagagatgaa gcaattgcag 840
cttgcagagc aggggctccc ttctctgtaa ttggaaacac aatcagccac ataactcatc 900
agaatggttt tcaagtctgt ccacattttg tgggacatgg aataggatct tactttcatg 960
gacatccaga aatttggcat catgcaaacg acagtgatct acccatggag gagggcatgg 1020
cattcactat agagccaatc atcacggagg gatcccctga atttaaagtc ctggaggatg 1080
catggactgt ggtctcccta gacaatcaaa ggtcggcgca gttcgagcac acggttctga 1140
tcacgtcgag gggcgcgcag atcctgacca aactacccca tgaggcctga ggagccgccc 1200
gaaggtcgcg gtgacctggt gccttttaa ataaattgct gaaatttggc tggagaactt 1260
ttagaagaaa cagggaaatg accggtggtg cggtaacctg cgtggctcct gatagcgttt 1320
ggaagaacgc gggggagact gaagagcaac tgggaactcg gatctgaagc cctgctgggg 1380
tcgcgcggct ttggaaaaac aaatcctggc cctggactcg gtttcccagc gcggtcaacg 1440
catctggagg ggactggagg aaaccccctt gttggaagag attccaagag aagcacggtt 1500
ttctctttcc cttgccctga ctgttggagt aaaaaacctc ttaaatccat tgtatcagag 1560
gtccttacct ctctgacagt tacaatgatc tttgtatctg aactttgcac gtctgccgaa 1620
aaatccgaac ctgttgactg ggatttttaa gaatccgttt ctccctttg tgtattccat 1680
attggccggc cccaaggatg ctcgcagaag ccagccccca accccagccc ttccgtatct 1740

ttcccctcca tcgcggcttt gcgatgaaag attagcccgc gaacagaggc attgattaca 1800
aacatgtcct tggccagtgg actctgggcc tggccattct tcaggtttct gtcaatccag 1860
aaacgcgact ttcctggacc cctgcggctc ttccttccca ccagctcagc atcacagccc 1920
atccagaggc caagtccaag aaggaataac agtaatgagg gaaccttccg agcaaaaacg 1980
caa 1983

<210> 29
<211> 1574
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474340CB1

<400> 29
gccagggcca agatggatct tctcctcgac atcagctaag cctggaggac tcttcccctc 60
agagaccatg gagagggaca gccacgggaa tgcatctcca gcaagaacac cttcagctgg 120
agcatctcca gcccaggcat ctccagctgg gacacctcca ggccgggcat ctccagccca 180
ggcatctcca gcccaggcat ctccagctgg gacacctccg ggccgggcat ctccagccca 240
ggcatctcca gctggtacac ctccaggccg ggcatctcca ggccgggcat ctccagccca 300
ggcatctcca gcccgggcat ctccggctct ggcatcactt tccaggtcct catccggcag 360
gtcatcatcc gccaggtcag cctcggtgac aacctcccca accagagtgt accttgttag 420
agcaacacca gtgggggctg tacccatccg atcatctcct gccaggtcag caccagcaac 480
cagggccacc agggagagcc caggtacgag cctgcccaag ttcacctggc gggagggcca 540
gaagcagcta ccgctcatcg ggtgcgtgct cctcctcatt gccctggtgg tttcgctcat 600
catcctcttc cagttctggc agggccacac agggatcagg tacaaggagc agagggagag 660
ctgtcccaag cacgctgttc gctgtgacgg ggtggtggac tgcaagctga agagtgacga 720
gctgggctgc gtgaggtttg actgggacaa gtctctgctt aaaatctact ctgggtcctc 780
ccatcagtgg cttcccatct gtagcagcaa ctggaatgac tcctactcag agaagacctg 840
ccagcagctg ggtttcgaga gtgctcaccg gacaaccgag gttgcccaca gggattttgc 900
caacagcttc tcaatcttga gatacaactc caccatccag gaaagcctcc acaggtctga 960
atgcccttcc cagcggtata tctcctcca gtgttcccac tgcggactga gggccatgac 1020
cgggcggatc gtgggagggg cgctggcctc ggatagcaag tggccttggc aagtgagtct 1080
gcacttcggc accacccaca tctgtggagg cacgctcatt gacgcccagt gggtgctcac 1140
tgccgcccac tgcttcttcg tgacccggga gaaggtcctg gagggctgga aggtgtacgc 1200
gggcaccagc aacctgcacc agttgcctga ggcagcctcc attgccgaga tcatcatcaa 1260
cagcaattac accgatgagg aggacgacta tgacatcgcc ctcatgcggc tgtccaagcc 1320
cctgaccctg tccggtgagg gaatctgcac tccccgctct cctgcccccc agccccagca 1380
ccctctgcag ccctcgcact tgtcagcatc tgtcaactca tatccgggcc ccaaagcttc 1440
tgcagggcag aagtcaaaga ctcttaaaga tccttacatg gaacacttct gttttataat 1500
tagggaaact gaagcccaag ggttataaat aagtttgctc caaatgacac atctcacatt 1560
acaaattgat gacg 1574

<210> 30
<211> 1173
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477287CB1

<400> 30
atggggccaa gactcattcc gtttctattt ttgtttgttt accctattct ctgcaggatc 60
attctgagga aaggcaagtc tatccgccag agaatggagg agcagggtgt actggagacg 120

tttctgaggg accacccaaa ggctgatcca attgccaagt attatttcaa taatgatgct 180
gttgcttatg agcccttcac caactacctg gattctttct actttgggga gatcagcact 240
gggacaccac cccaaaattt cctagtctct ttgatacggg ttcctccaat ctgtagcctg 300
ccctccatct actgccagag ccaagtctgc tccaatcaca acaggttcaa tcccagcctg 360
tcctccacct tcagaaacga tggacaaacc tatggactat cctatgggag tggcagcctg 420
agtgtgttcc tgggctatga cactgtgact gttcataaca tcgttgtcaa taaccaggag 480
tttggcctga gtgagaatga gcccagcgac ccctttact attcagactt tgacgggatc 540
ctgggaatgg cctacccaaa catggcagag gggaattccc ctacagtaat gcaggggatg 600
ctgcagcaga gccagcttac tcagcccgtc ttcagcttct acttcacctg ccagccaacc 660
cgccagtatt gtggagagct catccttgga ggtgtggacc ccaaccttta ttctggtcag 720
atcatctgga cccctgtcag cccggaactg tactggcaga ttgccatcga ggaatttgcc 780
atcggtaacc aggccactgg cttgtgctct gagggttgcc aggccattgt ggataccgag 840
accttcctgc tggcagttcc tcagcagtac atggcctcct tcctgcaggc aacaggaccc 900
cagcaggctc agaatggtga ctttgtggtc aactgcagct acatacagag catgcccacc 960
atcaccttca tcatcggcgg ggcccagttt cctctgcctc cctctgaata tgtcttcaat 1020
aacaatggct actgcaggct tggaactgag gccacctgcc tgccctcccg cagtgggcag 1080
cccctctgga ttctggggga tgtcttcctc aaggaatatt gctctgtcta tgacatggcc 1140
aacaacaggg tgggctttgc cttctctgcc tag 1173

<210> 31
<211> 6013
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2994162CB1

<400> 31
gcgggacctg gccgagatgg ggagcccaga cgccgcggcg gccgtgcgca aggacaggct 60
gcacccgagg caagtgaaat tattagagac cctgagcgaa tacgaaatcg tgtctcccat 120
ccgagtgaac gctctcggag aacccctttcc cacgaacgtc cacttcaaaa gaacgcgacg 180
gagcattaac tctgccactg acccctggcc tgccttcgcc tcctcctctt cctcctctac 240
ctcctcccag gcgcattacc gcctctctgc cttcggccag cagtttctat ttaatctcac 300
cgccaatgcc ggatttatcg ctccactgtt cactgtcacc ctcctcggga cgcccggggt 360
gaatcagacc aagttttatt ccgaagagga agcggaactc aagcactgtt tctacaaagg 420
ctatgtcaat accaactccg agcacacggc cgtcatcagc ctctgctcag gaatgctggg 480
cacattccgg tctcatgatg gggattattt tattgaacca ctacagtcta tggatgaaca 540
agaagatgaa gaggaacaaa acaaacccca catcatttat aggcgcagcg ccccccagag 600
agagccctca acaggaaggc atgcatgtga cacctcagaa cacaaaaata ggcacagtaa 660
agacaagaag aaaaccagag caagaaaatg gggagaaagg attaacctgg ctggtgacgt 720
agcagcatta aacagcggct tagcaacaga ggcattttct gcttatggta ataagacgga 780
caacacaaga gaaaagagga cccacagaag gacaaaacgt tttttatcct atccacggtt 840
tgtagaagtc ttggtggtgg cagacaacag aatggtttca taccatggag aaaaccttca 900
acactatatt ttaactttaa tgtcaattgt agcctctatc tataaagacc caagtattgg 960
aaatttaatt aatattgtta ttgtgaactt aattgtgatt cataatgaac aggatgggcc 1020
ttccatatct tttaatgctc agacaacatt aaaaaacttt tgccagtggc agcattcgaa 1080
gaacagtcca ggtggaatcc atcatgatac tgctgttctc ttaacaagac aggatatctg 1140
cagagctcac gacaaatgtg ataccttagg cctggctgaa ctgggaacca tttgtgatcc 1200
ctatagaagc tgttctatta gtgaagatag tggattgagt acagctttta cgatcgccca 1260
tgagctgggc catgtgttta acatgcctca tgatgacaac aacaaatgta aagaagaagg 1320
agttaagagt ccccagcatg tcatggctcc aacactgaac ttctacacca acccctggat 1380
gtggtcaaag tgtagtcgaa aatatatcac tgagttttta gacactggtt atggcgagtg 1440
tttgcttaac gaacctgaat ccagacccta ccctttgcct gtccaactgc caggcatcct 1500
ttacaacgtg aataaacaat gtgaattgat ttttggacca ggttctcagg tgtgcccata 1560
tatgatgcag tgcagacggc tctggtgcaa taacgtcaat ggagtacaca aaggctgccg 1620

gactcagcac acaccctggg ccgatgggac ggagtgcgag cctggaaagc actgcaagta 1680
tggattttgt gttcccaaag aaatggatgt ccccgtgaca gatggatcct ggggaagttg 1740
gagtcccttt ggaacctgct ccagaacatg tggagggggc atcaaaacag ccattcgaga 1800
gtgcaacaga ccagaaccaa aaaatggtgg aaaatactgt gtaggacgta gaatgaaatt 1860
taagtcctgc aacacggagc catgtctcaa gcagaagcga gacttccgag atgaacagtg 1920
tgctcacttt gacgggaagc attttaacat caacggtctg cttcccaatg tgcgctgggt 1980
ccctaaatac agtggaattc tgatgaagga ccggtgcaag ttgttctgca gagtggcagg 2040
gaacacagcc tactatcagc ttcgagacag agtgatagat ggaactcctt gtggccagga 2100
cacaaatgat atctgtgtcc agggcctttg ccggcaagct ggatgcgatc atgttttaaa 2160
ctcaaaagcc cggagagata aatgtggggt ttgtggtggc gataattctt catgcaaaac 2220
agtggcagga acatttaata cagtacatta tggttacaat actgtggtcc gaattccagc 2280
tggtgctacc aatattgatg tgcggcagca cagtttctca ggggaaacag acgatgacaa 2340
ctacttagct ttatcaagca gtaaaggtga attcttgcta aatggaaact ttgttgtcac 2400
aatggccaaa agggaaattc gcattgggaa tgctgtggta gagtacagtg ggtccgagac 2460
tgccgtagaa agaattaact caacagatcg cattgagcaa gaacttttgc ttcaggtttt 2520
gtcggtggga aagttgtaca accccgatgt acgctattct ttcaatattc caattgaaga 2580
taaacctcag cagttttact ggaacagtca tgggccatgg caagcatgca gtaaaccctg 2640
ccaaggggaa cggaaacgaa aacttgtttg caccagggaa tctgatcagc ttactgtttc 2700
tgatcaaaga tgcgatcggc tgccccagcc tggacacatt actgaaccct gtggtacaga 2760
ctgtgacctg aggtggcatg ttgccagcag gagtgaatgt agtgcccagt gtggcttggg 2820
ttaccgcaca ttggacatct actgtgccaa atatagcagg ctggatggga agactgagaa 2880
ggttgatgat ggtttttgca gcagccatcc caaaccaagc aaccgtgaaa aatgctcagg 2940
ggaatgtaac acgggtggct ggcgctattc tgcctggact gaatgttcaa aaagctgtga 3000
cggtgggacc cagaggagaa gggctatttg tgtcaatacc cgaaatgatg tactggatga 3060
cagcaaatgc acacatcaag agaaagttac cattcagagg tgcagtgagt tcccttgtcc 3120
acagtggaaa tctggagact ggtcagagtg cttggtcacc tgtggaaaag ggcataagca 3180
ccgccaggtc tggtgtcagt ttggtgaaga tcgattaaat gatagaatgt gtgaccctga 3240
gaccaagcca acatctatgc agacttgtca gcagccggaa tgtgcatcct ggcaggcggg 3300
tccctgggga cagtgcagtg tcacttgtgg acagggatac cagctaagag cagtgaaatg 3360
catcattggg acttatatgt cagtggtaga tgacaatgac tgtaatgcag caactagacc 3420
aactgatacc caggactgtg aattaccatc atgtcatcct cccccagctg ccccggaaac 3480
gaggagaagc acatacagtg caccaagaac ccagtggcga tttgggtctt ggaccccatg 3540
ctcagccact tgtgggaaag gtacccggat gagatacgtc agctgccgag atgagaatgg 3600
ctctgtggct gacgagagtg cctgtgctac cctgcctaga ccagtggcaa aggaagaatg 3660
ttctgtgaca ccctgtgggc aatggaaggc cttggactgg agctcttgct ctgtgacctg 3720
tgggcaaggt agggcaaccc ggcaagtgat gtgtgtcaac tacagtgacc acgtgatcga 3780
tcggagtgag tgtgaccagg attatatccc aaaaactgac caggactgtt ccatgtcacc 3840
atgccctcaa aggaccccag acagtggctt agctcagcac cccttccaaa atgaggacta 3900
tcgtccccgg agcgccagcc ccagccgcac ccatgtgctc ggtggaaacc agtggagaac 3960
tggcccctgg ggagcatgtt ccagtacctg tgctggcgga tcccagcggc gtgttgttgt 4020
atgtcaggat gaaaatggat acaccgcaaa cgactgtgtg gagagaataa aacctgatga 4080
gcaaagagcc tgtgaatccg gcccttgtcc tcagtgggct tatggcaact ggggagagtg 4140
cactaagctg tgtggtggag gcataagaac aagactggtg gtctgtcagc ggtccaacgg 4200
tgaacggttt ccagatttga gctgtgaaat tcttgataaa cctcccgatc gtgagcagtg 4260
taacacacat gcttgtccac acgacgctgc atggagtact ggcccttgga gctcgtgttc 4320
tgtctcttgt ggtcgagggc ataaacaacg aaatgtttac tgcatggcaa aagatggaag 4380
ccatttagaa agtgattact gtaagcacct ggctaagcca catgggcaca gaaagtgccg 4440
aggaggaaga tgccccaaat ggaaagctgg cgcttggagt cagtgctctg tgtcctgtgg 4500
ccgaggcgta cagcagaggc atgtgggctg tcagatcgga acacacaaaa tagccagaga 4560
gaccgagtgc aacccataca ccagaccgga gtcggaacgc gactgccaag gcccacggtg 4620
tcccctctac acttggaggg cagaggaatg gcaagaatgc accaagacct gcggcgaagg 4680
ctccaggtac cgcaaggtgg tgtgtgtgga tgacaacaaa aacgaggtgc atggggcacg 4740
ctgtgacgtg agcaagcggc cggtggaccg tgaaagctgt agtttgcaac cctgcgagta 4800
tgtctggatc acaggagaat ggtcagagtg ctcagtgacc tgtggaaaag gctacaaaca 4860
aaggcttgtc tcgtgcagcg agatttacac cgggaaggag aattatgaat acagctacca 4920
aaccaccatc aactgcccag gcacgcagcc ccccagtgtt caccctgtt acctgaggga 4980

ctgccctgtc tcggccacct ggagagttgg caactggggg agctgctcag tgtcttgtgg 5040
tgttggagtg atgcagagat ctgtgcaatg tttaaccaat gaggaccaac ccagccactt 5100
atgccacact gatctgaagc cagaagaacg aaaaacctgc cgtaatgtct ataactgtga 5160
gttaccccag aattgcaagg aggtaaaaag acttaaaggt gccagtgaag atggtgaata 5220
tttcctgatg attagaggaa agcttctgaa gatattctgt gcggggatgc actctgacca 5280
ccccaaagag tacgtgacac tggtgcatgg agactctgag aatttctccg aggtttatgg 5340
gcacaggtta cacaacccaa cagaatgtcc ctataacggg agccggcgcg atgactgcca 5400
atgtcggaag gattacacgg ccgctgggtt ttccagtttt cagaaaatca gaatagacct 5460
gaccagcatg cagataatca ccactgactt acagtttgca aggacaagcg aaggacatcc 5520
cgtccctttt gccacagccg gggattgcta cagcgctgcc aagtgcccac agggtcgttt 5580
tagcatcaac ctttatggaa ccggcttgtc tttaactgaa tctgccagat ggatatcaca 5640
agggaattat gctgtctctg acatcaagaa gtcgccggat ggtacccgag tcgtagggaa 5700
atgcggtggt tactgtggaa aatgcactcc atcctctggt actggcctgg aggtgcgagt 5760
tttatagcta aggtgctttg aagaggaagc cattatggat ggatgaagga tagtaatgca 5820
atacctccac cttaatttgg gtgcatgtgt atgtgtgtgt gtgtttgtgt gtgacttgta 5880
tgcttgtgtg tgtaaatgtg tgtacatata catatataca tatctacaca tacatatata 5940
cacatatatg tgtgtatgta gatatgtaga ctatcctaat gatgtaaagt ttaatattta 6000
tgtttgaaat tat 6013

<210> 32
<211> 1393
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3965293CB1

<400> 32
gcggccagag agctcgtcat ttgaagactc tctcggaagg gatagcgtct ttctgcaacc 60
tgcggtccca gcagacaaac cttgtgatcc tcgttccagt cgacatggag gacgactcac 120
tctacttggg aggtgagtgg cagttcaacc acttttcaaa actcacatct tctcggcccg 180
atgcagcttt tgctgaaatc cagcggactt ctctccctga gaagtcacca ctctcatgtg 240
agacccgtgt cgacctctgt gatgatttgg ctcctgtggc aagacagctt gctcccaggg 300
agaagcttcc tctgagtagc aggagacctg ctgcggtggg ggctgggctc cagaatatgg 360
gaaatacctg ctacgtgaac gcttccttgc agtgcctgac atacacaccg ccccttgcca 420
actacatgct gtcccgggag cactctcaaa cgtgtcatcg tcacaagggc tgcatgctct 480
gtactatgca agctcacatc acacgggccc tccacaatcc tggccacgtc atccagccct 540
cacaggcatt ggctgctggc ttccatagag gcaagcagga agatgcccat gaatttctca 600
tgttcactgt ggatgccatg aaaaaggcat gccttcccgg gcacaagcag gtagatcatc 660
actctaagga caccaccctc atccaccaaa tatttggagg ctactggaga tctcaaatca 720
agtgtctcca ctgccacggc atttcagaca cttttgaccc ttacctggac atcgccctgg 780
atatccaggc agctcagagt gtccagcaag ctttggaaca gttggtgaag cccgaagaac 840
tcaatggaga gaatgcctat cattgtggtg tttgtctcca gagggcgccg gcctccaaga 900
cgttaacttt acacacctct gccaaggtcc tcatccttgt attgaagaga ttctccgatg 960
tcacaggcaa cctcgagccg aattcggctc gagctcgagc cgaaagatcc caatgttcta 1020
cctctccctg tccctcttgt aggggatagg gaggcagaga gagccagccc ctaccctcag 1080
agtatctgga cctcagagac catgttgtgc caggggtggt cccacctaaa gatgctagcc 1140
cctctccagg tgggcataag gagtaacaga tggcaaaacc acaaactatt ttgatggact 1200
gtgctgcagt atcaccagaa gacattaggg ggcagtaggc ccccacacaa aaccttcagg 1260
cttgaatttt aaaggggagg actttctgcc aacttttctt gtatgccttg ggaaagccag 1320
ttgccctgaa cccagcagac accatggaat gtcctttgca cgcattaaat ggtacagaac 1380
tgaaaaaaaa aaa 1393

<210> 33
<211> 1993

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4948403CB1

<400> 33
cccaaaggaa gcagcaccca gtgaacccct gccgtgagtc agcacgaggg aggcccagcc 60
ctttctagag gagcctgatt aaagatcagg ctcagctgct gctgctgctg ctgctgcttg 120
tcccaagacc aagtcgtaat agcaacttcc cttcctcagc tgcctgaact ttttttttcc 180
cttgtagctg gagagaagtg tcacattttg ctcactctca accttcctcg cccacccccc 240
tcccggagaa cctgtgcggt gtgtagaggg tgctgtgagc cacctccagc ctcgggtggc 300
tgcttaagta actttcaact cctctcttct taacactatg aagtgtctcg ggaagcgcag 360
gggccaggca gctgctttcc tgcctctttg ctggctcttt ttgaagattc tgcaaccggg 420
gcacagccac ctttataaca accgctatgc tggtgataaa gtgataagat ttattcccaa 480
aacagaagag gaagcatatg cactgaagaa aatatcctat caacttaagg tggacctgtg 540
gcagcccagc agtatctcct atgtatcaga gggaacagtt actgatgtcc atatccccca 600
aaatggttcc cgagccctgt tagccttctt acaggaagcc aacatccagt acaaggtcct 660
catagaagat cttcagaaaa cactggagaa gggaagcagc ttgcacaccc agagaaaccg 720
aagatccctc tctggatata attatgaagt ttatcactcc ttagaagaaa ttcaaaattg 780
gatgcatcat ctgaataaaa ctcactcagg cctcattcac atgttctcta ttggaagatc 840
atatgaggga agatctcttt ttattttaaa gctgggcaga cgatcacgac tcaaaagagc 900
tgtttggata gactgtggta ttcatgcaag agaatggatt ggtcctgcct tttgtcagtg 960
gtttgtaaaa gaagctcttc taacatataa gagtgaccca gccatgagaa aaatgttgaa 1020
tcatctatat ttctatatca tgcctgtgtt taacgtcgat ggataccatt ttagttggac 1080
caatgatcga ttttggagaa aaacaaggtc aaggaactca aggtttcgct gccgtggagt 1140
ggatgccaat agaaactgga aagtgaagtg gtgtgatgaa ggagcttcta tgcacccttg 1200
tgatgacaca tactgtggcc cttttccaga atctgagccg gaagtgaagg ctgtagctaa 1260
cttccttcga aaacacagaa agcacattag ggcttatctc tcctttcatg catatgctca 1320
gatgttactg tatccctatt cttacaaata tgcaacaatt cccaatttta gatgtgtgga 1380
atctgcagct tataaagctg tgaatgcact tcagtcagta tacggggtac gatacagata 1440
tggaccagcc tccacaacgt tgtatgtgag ctctggtagc tcaatggatt gggcctacaa 1500
aaatggaata ccttatgcat ttgctttcga actacgtgac actggatatt ttggattttt 1560
actcccagag atgctcatca aacccacctg tacagaaact atgctggctg tgaaaaatat 1620
cacaatgcac ctgctaaaga aatgtccctg agacagccca aggctcaggt caactgccat 1680
aggattctga gcaaggccta cttggccctg gatagaaatt gttttcaaag agaagggcag 1740
ctgcttagag tgaacatgtc tatggacttt aaaaagaccc cacgcaattt gactttgtgg 1800
caatagaaaa cagtaaaaaa cagggcatag cctagtttgt tataagaaaa agcatccatt 1860
ttctatcctt ttagagtctt atttgattat ggtgggaggg aatgttttca aatttcccat 1920
ttctcaagaa atgttcatat taattgagga tttcccttca ataaatctca tgtcctcagt 1980
taggaaaaaa aaa 1993

<210> 34
<211> 2318
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473165CB1

<400> 34
cggcagccac tcctgagtga gcaaaggttc ctccgcggtg ctctcccgtc cagagccctg 60
ctgatgggga agtccgaggg ccagtgggga tggtggagag cgccggccgt gcagggcaga 120
agcgcccggg gttcctggag ggggggctgc tgctgctgct gctgctggtg accgctgccc 180

tggtggcctt gggtgtcctc tacgccgacc gcagagggat cccagaggcc caagaggtga 240
gcgaggtctg caccaccect ggctgcgtga tagcagccgc caggatcctc cagaacatgg 300
acccgaccac ggaaccgtgt gacgacttct accagtttgc atgcggaggc tggctgcggc 360
gccacgtgat ccctgagacc aactcaagat acagcatctt tgacgtcctc cgcgacgagc 420
tggaggtcat cctcaaagcg gtgctggaga attcgactgc caaggaccgg ccggctgtgg 480
agaaggccag gacgctgtac cgctcctgca tgaaccagag tgtgatagag aagcgaggct 540
ctcagcccct gctggacatc ttggaggtgg tgggaggctg gccggtggcg atggacaggt 600
ggaacgagac cgtaggactc gagtgggagc tggagcggca gctggcgctg atgaactcac 660
agttcaacag gcgcgtcctc atcgacctct tcatctggaa cgacgaccag aactccagcc 720
ggcacatcat ctacatagac cagcccacct tgggcatgcc ctcccgagag tactacttca 780
acggcggcag caaccggaag gtgcgggaag cctacctgca gttcatggtg tcagtggcca 840
cgttgctgcg ggaggatgca aacctgccca gggacagctg cctggtgcag gaggacatgg 900
tgcaggtgct ggagctggag acacagctgg ccaaggccac ggtaccccag gaggagagac 960
acgacgtcat cgccttgtac caccggatgg gactggagga gctgcaaagc caatttggcc 1020
tgaagggatt taactggact ctgttcatac aaactgtgct atcctctgtc aaaatcaagc 1080
tgctgccaga tgaggaagtg gtggtctatg gcatccccta cctgcagaac cttgaaaaca 1140
tcatcgacac ctactcagcc aggaccatac agaactacct ggtctggcgc ctggtgctgg 1200
accgcattgg tagcctaagc cagagattca aggacacacg agtgaactac cgcaaggcgc 1260
tgtttggcac aatggtggag gaggtgcgct ggcgtgaatg tgtgggctac gtcaacagca 1320
acatggagaa cgccgtgggc tccctctacg tcagggaggc gttccctgga gacagcaaga 1380
gcatggtgga actcattgac aaggtgcgga cagtgtttgt ggagacgctg gacgagctgg 1440
gctggatgga cgaggagtcc aagaagaagg cgcaggagaa ggccatgagc atccgggagc 1500
agatcgggca ccctgactac atcctggagg agatgaacag gcgcctggac gaggagtact 1560
ccaatgtgaa cttctcagag gacctgtact ttgagaacag tctgcagaac ctcaaggtgg 1620
gcgcccagcg gagcctcagg aagcttcggg aaaaggtgga cccaaatctg atcatcgggg 1680
cggcggtggt caatgcgttc tactccccaa accgaaacca gattgtattc cctgccggga 1740
tcctccagcc ccccttcttc agcaaggagc agccacaggc cttgaacttt ggaggcattg 1800
ggatggtgat cgggcacgag atcacgcacg gctttgacga caatggtcgg aacttcgaca 1860
agaatggcaa catgatggat tggtggagta acttctccac ccagcacttc cgggagcagt 1920
cagagtgcat gatctaccag tacggcaact actcctggga cctggcagac gaacagaacg 1980
tgaacggatt caacaccctt ggggaaaaca ttgctgacaa cggaggggtg cggcaagcct 2040
ataaggccta cctcaagtgg atggcagagg gtggcaagga ccagcagctg cccggcctgg 2100
atctcaccca tgagcagctc ttcttcatca actatgccca ggtgtggtgc gggtcctacc 2160
ggcccgagtt cgccatccaa tccatcaaga cagacgtcca cagtcccctg aagtacaggg 2220
tactggggtc gctgcagaac ctggccgcct tcgcagacac gttccactgt gcccggggca 2280
cccccatgca ccccaaggag cgatgccgcg tgtggtag 2318

<210> 35
<211> 1931
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476667CB1

<400> 35
cccttatatc atcgtcttcg ccatctacaa atgaaatgtt caccctaact accaatgggg 60
acctaccccg accaatattc atccccaatg gaatgccaaa cactgttgtg ccatgtggaa 120
ctgagaagaa cttcacaaat ggaatggtta atggtcacat gccatctctt cctgacagcc 180
cctttacagg ttacatcatt gcagtccacc gaaaaatgat gaggacagaa ctgtatttcc 240
tgtcatctca gaagaatcgc cccagcctct ttggaatgcc attgattgtt ccatgtactg 300
tgcatacccg gaagaaagac ctatatgatg cggtttggat tcaagtatcc cggttagcga 360
gcccactccc acctcaggaa gctagtaatc atgcccagga ttgtgacgac agtatgggct 420
atcaatatcc attcactcta cgagttgtgc agaaagatgg gaactcctgt gcttggtgcc 480
catggtatag attttgcaga ggctgtaaaa ttgattgtgg ggaagacaga gctttcattg 540

gaaatgccta tatcgctgtg gattgggatc ccacagccct tcaccttcgc tatcaaacat 600
cccaggaaag ggttgtagat gagcatgaga gtgtggagca gagtcggcga gcgcaagccg 660
agcccatcaa cctggacagc tgtctccgtg ctttcaccag tgaggaagag ctaggggaaa 720
atgagatgta ctactgttcc aagtgtaaga cccactgctt agcaacaaag aagctggatc 780
tctggaggct tccacccatc ctgattattc accttaagcg atttcaattt gtaaatggtc 840
ggtggataaa atcacagaaa attgtcaaat ttcctcggga aagttttgat ccaagtgctt 900
ttttggtacc aagagacccg gctctctgcc agcataaacc actcacaccc caggggggatg 960
agctctctga gcccaggatt ctggcaaggg aggtgaagaa agtggatgcg cagagttcgg 1020
ctggggaaga ggacgtgctc ctgagcaaaa gcccatcctc actcagcgct aacatcatca 1080
gcagcccgaa aggttctcct tcttcatcaa gaaaaagtgg aaccagctgt ccctccagca 1140
aaaacagcag ccctaatagc agcccacgga ctttggggag gagcaaaggg aggctccggc 1200
tgccccagat tggcagcaaa aataaactgt caagtagtaa agagaacttg gatgccagca 1260
aagaaaatgg ggctgggcag atatgtgagc tggctgacgc cttgagtcga gggcatgtgc 1320
tgggggcag ccaaccagag ttggtcactc ctcaggacca tgaggtagct ttggccaatg 1380
gattccttta tgagcatgaa gcatgtggca atggctacag caatggtcag cttggaaacc 1440
acagtgaaga agacagcact gatgaccaaa gagaagatac tcgtattaag cctatttata 1500
atctatatgc aatttcgtgc cattcaggaa ttctgggtgg gggccattac gtcacttatg 1560
ccaaaaaccc aaactgcaag tggtactgtt acaatgacag cagctgtaag gaacttcacc 1620
cggatgaaat tgacaccgac tctgcctaca ttctttttcta tgagcagcag gggatagact 1680
atgcacaatt tctgccaaag actgatggca aaaagatggc agacacaagc agtatggatg 1740
aagactttga gtctgattac aaaaagtact gtgtgttaca gtaaagctac cactctggct 1800
gctagacagc ttggcggtga gggagatgac tccttgtagc tgacatttgg caaaagcgtc 1860
actgaaaggc aagctaaatg tagttatttt atcctgtggc cctgaagcaa aaaataaaaa 1920
ttcgaattaa g 1931

<210> 36
<211> 1218
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7479166CB1

<400> 36
atgctcagcc ccccgcagcc caggacccct gactgtaggc tccaggcctc cctggaagcc 60
ctggccacgc tcgccccgca gccctcagac tggctgtgct tcgcggatct tggctggttc 120
gaggctgatg gagctgccca ctccatgggc ctgggcagca gcttgaagtg ggcgtgggcc 180
aagccctctg ggatgcccgt cccagagaat gacctggtgg gcattgtggg gggccacaat 240
gccccccggg ggaagtggcc gtggcaggtc agcctgaggg tctacagcta ccactgggcc 300
tcctgggcgc acatctgtgg gggctccctc atccaccccc agtgggtgct gactgctgcc 360
cactgcattt tctggaagga caccgacccg tccatctacc ggatccacgc tggggacgtg 420
tatctctacg ggggccgggg gctgctgaac gtcagccgga tcatcgtcca ccccaactat 480
gtcactgcgg ggctgggtgc ggatgtggcc ctgctccagc tgccggggtc acctctctcc 540
ccagagtcgc tgccgccgcc ctaccgcctg cagcaggcga gtgtgcaggt gctggagaac 600
gccgtctgtg agcagcccta ccgcaacgcc tcagggcaca ctggcgaccg gcagctcatc 660
ctggatgaca tgctgtgtgc cggcagcgag ggccgagact cctgctacgg tgactccggc 720
ggccctctgg tctgcaggct gcgggggtcc tggcgcctgg tgggggtggt cagctggggc 780
tacggctgta ccctgcggga ctttcccggc gtctacaccc acgtccagat ctacgtgctc 840
tggatcctgc agcaagtcgg ggagttgccc tgagcaggct gggctgggct cccacctggg 900
tcggctgagg agggaccagg accttcctcc tcccagcgat ctccgcttcg gcctccgctg 960
caggccaccg tcttgagccc ggcttctctg gctcctcagc gcccaggacc tccctgatgc 1020
cggggtgggg aaggggccgg ggaagggagg gtgggggcct cgctgcgtct ctgtctgatt 1080
aaagagcaag agcagagtgt gtggcgtctc tgtgggatgg atttgcattc caagctgcag 1140
ccaggtgcgg tttgctcagc cacctcctgt tggaggcctc cacattttgg ctatggtaat 1200
aaagatgctg agaaaatt 1218

<210> 37
<211> 2679
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3671788CB1

<400> 37
caattaatat taacgaggga aggctcctca ttgcctaaag accccactgg ggctccaatg 60
gaagagaggc cccgcccccg tgactcagag gttaaagggc ctggtgccgg cttgtgaggc 120
cagtgtccag atggcatcca gcagtgggag ggtcaccatc cagctcgtgg atgaggaggc 180
tggggtcgga gccgggcgcc tgcagctttt tcggggccag agctatgagg caattcgggc 240
agcctgcctg gattcgggga tcctgttccg cgacccttac ttccctgctg gccctgatgc 300
ccttggctat gaccagctgg ggccggactc ggagaaggcc aaaggcgtga aatggatgag 360
gccccatgag ttctgtgctg agccgaagtt catctgtgaa gacatgagcc gcacagacgt 420
gtgtcagggg agcctgggta actgctggtt ccttgcagcc gccgcctccc ttactctgta 480
tccccggctc ctgcgccggg tggtccctcc tggacaggat ttccagcatg gctacgcagg 540
cgtcttccac ttccagctct ggcagtttgg ccgctggatg gacgtcgtgg tggatgacag 600
gctgcccgtg cgtgagggga agctgatgtt cgtgcgctcg gaacagcgga atgagttctg 660
ggccccactc ctggagaagg cctacgccaa gctccacggc tcctatgagg tgatgcgggg 720
cggccacatg aatgaggctt ttgtggattt cacaggcggc gtgggcgagg tgctctatct 780
gagacaaaac agcatggggc tgttctctgc cctgcgccat gccctggcca aggagtccct 840
cgtgggcgcc actgccctga gtgatcgggg tgagtaccgc acagaagagg gcctggtaaa 900
gggacacgcg tattccatca cgggcacaca caaggtgttc ctgggcttca ccaaggtgcg 960
gctgctgcgg ctgcggaacc catggggctg cgtggagtgg acgggggcct ggagcgacag 1020
ctgcccacgc tgggacacac tccccaccga gtgccgcgat gccctgctgg tgaaaaagga 1080
ggatggcgag ttctggatgg agctgcggga cttcctcctc catttcgaca ccgtgcagat 1140
ctgctcgctg agcccggagg tgctgggccc cagcccggag gggggcggct ggcacgtcca 1200
caccttccaa ggccgctggg tgcgtggctt caactccggc gggagccagc ctaatgctga 1260
aaccttctgg accaatcctc agttccgttt aacgctgctg gagcctgatg aggaggatga 1320
cgaggatgag gaagggccct gggggggctg gggggctgca ggggcacggg gcccagcgcg 1380
gggggccgc acgcccaagt gcacggtcct tctgtccctc atccagcgca accggcggcg 1440
cctgagagcc aagggcctca cttacctcac cgttggcttc cacgtgttcc aggcagaggg 1500
ctccacaggc acagacaacg agcggacaca cggcttcacc ggacacagag gagcacagct 1560
cgccggtcac acacacggcc cacaagaggc gagcaaaaga tacacgcaga acagcgctga 1620
ggtagcccca gatagggaag cggacgacga cgggggacag gggttcggcg acgggccatg 1680
ggagatcgac gacgtgatca gcgcagacct gcagtctctc cagggccccc acctgccccc 1740
ggagctgggg ttggagcagc tgtttcagga gctggctgga gaggaggaag aactcaatgc 1800
ctctcagctc caggccttac taagcattgc cctggagcct gccagggccc atacctccac 1860
ccccagagag atcgggctca ggacctgtga gcagctgctg cagtgtttcg ggcatgggca 1920
aagcctggcc ttacaccact tccagcagct ctggggctac ctcctggagt ggcaggccat 1980
attcaacaag ttcgatgagg acacctctgg aaccatgaac tcctacgagc tgaggctggc 2040
actgaatgca gcaggcttcc acctgaacaa ccagctgacc cagaccctca ccagccgcta 2100
ccgggatagc cgtctgcgtg tggacttcga gcggttcgtg tcctgtgtgg cccacctcac 2160
ctgcatcttc tgccactgca gccagcacct ggatgggggt gaggggtca tctgcctgac 2220
ccacagacag tggatggagg tggccacctt ctcctaggat ctccggatgg gcgcacctgc 2280
tgctcagggc agggttgctg agcaagacca cctccctagg ccttgcctgg catgggtgcc 2340
actctctctg gcatccacct gtctgggggct agtctctggc cctcactgct cacggccggg 2400
tgaccactct ggcctgcgta ctcctcactc agaaacaaga acagcgacag cccttctcga 2460
gcagatgaca cgagctagtc cacgttgaca gcttaagaca gtgctagctc tgccctggct 2520
ctcctagaag gtggaggaca gacacaggag aaataaaaaa agatgatgct gcaggaatcc 2580
ttcttaaaaa tattacatgt tttattatcc tgtccccaga gggtggttta tccagaaacc 2640
aagaaaaaaa atcaatcaga ataaactcaa aaaaaaaaa 2679

<210> 38
<211> 2632
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7479181CB1

<400> 38
gggagagcct ggcgagctga aacccgagct cccgctcagc tggggctcgg ggaggtccct 60
gtaaaacccg cctgcccccg gcctccctgg gtccctcctc tccctcccca gtagacgctc 120
gggcaccagc cgcggcaagg atggagctgg gttgctggac gcagttgggg ctcacttttc 180
ttcagctcct tctcatctcg tccttgccaa gagagtacac agtcattaat gaagcctgcc 240
ctggagcaga gtggaatatc atgtgtcggg agtgctgtga atatgatcag attgagtgcg 300
tctgccccgg aaagagggaa gtcgtgggtt ataccatccc ttgctgcagg aatgaggaga 360
atgagtgtga ctcctgcctg atccacccag gttgtaccat ctttgaaaac tgcaagagct 420
gccgaaatgg ctcatggggg ggtaccttgg atgacttcta tgtgaagggg ttctactgtg 480
cagagtgccg agcaggctgg tacggaggag actgcatgcg atgtggccag gttctgcgag 540
ccccaaaggg tcagattttg ttggaaagct atcccctaaa tgctcactgt gaatggacca 600
ttcatgctaa acctgggttt gtcatccaac taagatttgt catgttgagc ctggagtttg 660
actacatgtg ccagtatgac tatgttgagg ttcgtgatgg agacaaccgc gatggccaga 720
tcatcaagcg tgtctgtggc aacgagcggc cagctcctat ccagagcata ggatcctcac 780
tccacgtcct cttccactcc gatggctcca agaattttga cggtttccat gccatttatg 840
aggagatcac agcatgctcc tcatcccctt gtttccatga cggcacgtgc gtccttgaca 900
aggctggatc ttacaagtgt gcctgcttgg caggctatac tgggcagcgc tgtgaaaatc 960
cctgccgaga accaaagatt tcagacctgg tgagaaggag agttcttccg atgcaggttc 1020
agtcaaggga gacaccatta caccagctat actcagcggc cttcagcaag cagaaactgc 1080
agagtgcccc taccaagaag ccagcccttc cctttggaga tctgcccatg ggataccaac 1140
atctgcatac ccagctccag tatgagtgca tctcaccctt ctaccgccgc ctgggcagca 1200
gcaggaggac atgtctgagg actgggaagt ggagtgggcg ggcaccatcc tgcatcccta 1260
tctgcgggaa aattgagaac atcactgctc caaagaccca agggttgcgc tggccgtggc 1320
aggcagccat ctacaggagg accagcgggg tgcatgacgg cagcctacac aagggagcgt 1380
ggttcctagt ctgcagcggt gccctggtga atgagcgcac tgtggtggtg gctgcccact 1440
gtgttactga cctggggaag gtcaccatga tcaagacagc agacctgaaa gttgttttgg 1500
ggaaattcta ccgggatgat gaccgggatg agaagaccat ccagagccta cagatttctg 1560
ctatcattct gcatcccaac tatgacccca tcctgcttga tgctgacatc gccatcctga 1620
agctcctaga caaggcccgt atcagcaccc gagtccagcc catctgcctc gctgccagtc 1680
gggatctcag cacttccttc caggagtccc acatcactgt ggctggctgg aatgtcctgg 1740
cagacgtgag gagccctggc ttcaagaacg acacactgcg ctctggggtg gtcagtgtgg 1800
tggactcgct gctgtgtgag gagcagcatg aggaccatgg catcccagtg agtgtcactg 1860
ataacatgtt ctgtgccagc tgggaaccca ctgccccttc tgatatctgc actgcagaga 1920
caggaggcat cgcggctgtg tccttcccgg gacgagcatc tcctgagcca cgctggcatc 1980
tgatgggact ggtcagctgg agctatgata aaacatgcag ccacaggctc tccactgcct 2040
tcaccaaggt gctgcctttt aaagactgga ttgaaagaaa tatgaaatga accatgctca 2100
tgcactcctt gagaagtgtt tctgtatatc cgtctgtacg tgtgtcattg cgtgaagcag 2160
tgtgggcctg aagtgtgatt tggcctgtga acttggctgt gccagggctt ctgacttcag 2220
ggacaaaact cagtgaaggg tgagtagacc tccattgctg gtaggctgat gccgcgtcca 2280
ctactaggac agccaattgg aagatgccag ggcttgcaag aagtaagttt cttcaaagaa 2340
gaccatatac aaaacctctc cactccactg acctggtggt cttccccaac tttcagttat 2400
acgaatgcca tcagcttgac cagggaagat ctgggcttca tgaggcccct tttgaggctc 2460
tcaagttcta gagagctgcc tgtgggacag cccagggcag cagagctggg atgtggtgca 2520
tgcctttgtg tacatggcca cagtacagtc tggtcctttt ccttccccat ctcttgtaca 2580
cattttaata aaataagggt tggcttctga actacaaaaa aaaaaaaaaa aa 2632

<210> 39

<211> 2757
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6621372CB1

<400> 39
atgccagggg gcgcaggcgc cgccaggctc tgcttgctgg cgtttgccct gcagcccctc 60
cggccgcggg cggcgcggga gcctggatgg acaagaggaa gtgaggaagg cagccccaag 120
ctgcagcatg aacttatcat acctcagtgg aagacttcag aaagccccgt gagagaaaag 180
catccactca aagctgagct cagggtaatg gctgaggggc gagaactgat cctggacctg 240
gagaagaatg agcaactttt tgctccttcc tacacagaaa cccattatac ttcaagtggt 300
aaccctcaaa ccaccacacg gaaattggag gatcactgct tttaccacgg cacggtgagg 360
gagacagaac tgtccagcgt cacgctcagc acttgccgag gaattagagg actgattacg 420
gtgagcagca acctcagcta cgtcatcgag cccctccctg acagcaaggg ccaacacctt 480
atttacagat ctgaacatct caagccgccc ccgggaaact gtgggttcga gcactccaag 540
cccaccacca gggactgggc tcttcagttt acacaacaga ccaagaagcg acctcgcagg 600
atgaaaaggg aagatttaaa ctccatgaag tatgtggagc tttacctcgt ggctgattat 660
ttagagtttc agaagaatcg acgagaccag gacgccacca aacacaagct catagagatc 720
gccaactatg ttgataagtt ttaccgatcc ttgaacatcc ggattgctct cgtgggcttg 780
gaagtgtgga cccacgggaa catgtgtgaa gtttcagaga atccatattc taccctctgg 840
tcctttctca gttggaggcg caagctgctt gcccagaagt accatgacaa cgcccaatta 900
atcacgggca tgtccttcca cggcaccacc atcggcctgg ccccctcat ggccatgtgc 960
tctgtgtacc agtctggagg agtcaacatg gaccactccg agaatgccat tggcgtggct 1020
gccaccatgg cccacgagat gggccacaac tttggcatga cccatgattc tgcagattgc 1080
tgctcggcca gtgcggctga tggtgggtgc atcatggcag ctgccactgg gcacccctttt 1140
cccaaagtgt tcaatggatg caacaggagg gagctggaca ggtatctgca gtcaggtggt 1200
ggaatgtgtc tctccaacat gccagacacc aggatgttgt atggaggccg gaggtgtggg 1260
aacgggtatc tggaagatgg ggaagagtgt gactgtggag aagaagagga atgtaacaac 1320
ccctgctgca atgcctctaa ttgtaccctg aggccggggg cggagtgtgc tcacggctcc 1380
tgctgccacc agtgtaagct gttggctcct gggaccctgt gccgcgagca ggccaggcag 1440
tgtgacctcc cggagttctg tacgggcaag tctccccact gccctaccaa cttctaccag 1500
atggatggta ccccctgtga gggcggccag gcctactgct acaacggcat gtgcctcacc 1560
taccaggagc agtgccagca gctgtgggga cccggagccc gacctgcccc tgacctctgc 1620
ttcgagaagg tgaatgtggc aggagacacc tttggaaact gtggaaagga catgaatggt 1680
gaacacagga agtgcaacat gagagatgcg aagtgtggga agatccagtg tcagagctct 1740
gaggcccggc ccctggagtc caacgcggtg cccattgaca ccactatcat catgaatggg 1800
aggcagatcc agtgccgggg cacccacgtc taccgaggtc ctgaggagga gggtgacatg 1860
ctggacccag ggctggtgat gactggaacc aagtgtggct acaaccatat ttgctttgag 1920
gggcagtgca ggaacacctc cttctttgaa actgaaggct gtgggaagaa gtgcaatggc 1980
catggggtct gtaacaacaa ccagaactgc cactgcctgc cgggctgggc cccgcccttc 2040
tgcaacacac cgggccacgg gggcagtatc gacagtgggc ctatgccccc tgagagtgtg 2100
ggtcctgtgg tagctggagt gttggtggcc atcttggtgc tggcggtcct catgctgatg 2160
tactactgct gcagacagaa caacaaacta ggccaactca agccctcagc tctcccttcc 2220
aagctgaggc aacagttcag ttgtcccttc agggtttctc agaacagcgg gactggtcat 2280
gccaacccaa ctttcaagct gcagacgccc cagggcaagc gaaaggtgat caacactccg 2340
gaaatcctgc ggaagccctc ccagcctcct ccccggcccc ctccagatta tctgcgtggt 2400
gggtccccac ctgcaccact gccagctcac ctgagcaggg ctgctaggaa ctccccaggg 2460
cccgggtctc aaatagagag gacggagtcg tccaggaggc ctcctccaag ccggccaatt 2520
ccccccgcac caaattgcat cgtttcccag gacttctcca ggcctcggcc gccccagaag 2580
gcactcccgg caaacccagt gccaggccgc aggagcctcc ccaggccagg aggtgcatcc 2640
ccactgcggc cccctggtgc tggccctcag cagtcccggc ctctggcagc acttgccccа 2700
aagtttccag aatacagatc acagagggct ggagggatga ttagctcgaa aatctag 2757

<210> 40
<211> 1892
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4847254CB1

<400> 40
ttcttcaggt tgttcggccg ttgttctctg tgtgctccgt tctgggggtg tctttgtagt 60
cttggcctct gtttttcatg tgttgcgctc tcgcctcgcg gcctcccttt cccgcgcccc 120
gtcgtcgtag tcctgctctg cctcttgctt tgtcttcttc tgtatctttc tgcttcgttt 180
cctgtcttcg ttctctcatg tttctttcgt gctgccgtct tctcgctcgc gtcttctgtc 240
tctcgttctc gtcatgtttc tcttctcgtc cccgtccctg tctcctgtct tcctcttgta 300
tctcctcctc ctctgcctct cctagaatct ccctcgccct cgccccgctc ctccatgaac 360
tcgcacggca ccgtccccgc ctctccagaa tcccccgtcc ccgcccccag aatctccccg 420
ccccgccccc agaacccccg ccccgccccc agaaccccgg ccccgccccc agaaccccgg 480
ccccgcgagg atgagcccag ggctccacgg tccctaccta gaccccacgc gatccctcac 540
ctgagacccc gtcccacaca gccccagctg gggcaaacag ccccctcccc acttcccatc 600
tgtaatttgc agggagatcg acgacgtgat cagcgcagac ctgcagtctc tccaggtggg 660
gactgttcct ggaggggcgg catggggcgg ggatcttggc cagcgctaaa cttccgccat 720
gcggcagggc ccctacctgc ccctggagct ggggttggag cagctgtttc aggagctggc 780
tggagaggag gaagaactca atgcctctca gctccaggcc ttactaagca ttgccctgga 840
gcctgccagg gcccatacct ccacccccag agagatcggg ctcaggacct gtgagcagct 900
gctgcagtgt ttcggggtac atgggggcca gtgcctgggt gagggaggga gtggggaagg 960
ggacgttggg gtctctcctc cccttctgga gagattgacc ttaaccagat gccccccgacc 1020
cccaacacag catgggcaaa gcctggcctt acaccacttc cagcagctct ggggctacct 1080
cctggagtgg caggccatat ttaacaagtt cgatgaggac acctctggaa ccatgaactc 1140
ctacgagctg aggctggcac tgaatgcagc aggcttccac ctgaacaacc agctgaccca 1200
gaccctcacc agccgctacc gggatagccg tctgcgtgtg gacttcgagc ggttcgtgtc 1260
ctgtgtggcc cacctcacct gcatcttctg ccactgcagc cagcacctgg atgggggtga 1320
gggggtcatc tgcctgaccc acagacagtg gatggaggtg gccaccttct cctaggatct 1380
ccggatgggc gcacctgctg ctcagggcag ggttgctgag caagaccacc tccctaggcc 1440
ttgcctggca tgggtgccac tctctctggc atccacctgt ctgggggctag tctctggccc 1500
tcactgctca cggccgggtg accactctgg cctgcgtact cctcactcag aaacaagaac 1560
agcgacagcc ccttctcgag cagatgacac gagctagtcc acgttgacag cttaagacag 1620
gtgctagctc tgcctggctc tcctagaagg tggaggacag acacgggaga aatacacaaa 1680
gatgaatgtt gccaggaatt ccttctttaa aatttcacca tgtgttatta tcctgtcccc 1740
agagggtggt ttatccagaa accaggaaaa aatcatccga taactccaaa aaaaaaaggg 1800
ggccgcgata tgggccggcg acgggaataa ccggaccgac tgggcggggg gagatcaatc 1860
agcttggacc gcccgggggg cggccaatcc tg 1892

<210> 41
<211> 3172
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5776350CB1

<220>
<221> unsure
<222> 3158
<223> a, t, c, g, or other

<400> 41
atgaagctgg agccattaca agagcgtgag cccgcgccgg aggagaactt gacgtggagc 60
agcagcggcg gcgacgagaa ggtgctccct tcaatccccc ttcgctgtca cagcagctcc 120
tcgcccgttt gcccgcgccg caagccccgc cctcggcccc agccccgggc ccgctcccgc 180
agccagcctg ggctctcggc cccacccccg cctccagccc ggcccccgcc cccgccgcca 240
cccccgcccc cacccgcacc gcggcccagg gcctggcgtg gatcccggcg cagatcccgg 300
cctgggtcca ggcctcagac acggagaagc tgctctggtg acctagacgg gtcgggggat 360
cctggcggct taggggactg gttgctggaa gtcgagtttg gtcagggtcc cacaggctgc 420
tctcatgtgg agagctttaa agtaggtaag aactggcaga agaacctgag gttgatctac 480
cagcgtttcg tttggagtgg gaccccagag actaggaaac gtaaagcaaa gtcatgcatc 540
tgtcacgtat gtagtaccca tatgaacaga ctccactctt gtctctcctg tgtcttttt 600
ggctgcttca ctgagaaaca tattcacaaa catgcagaaa caaagcagca ccatttagct 660
gtagaccttt atcatggggt catatattgc ttcatgtgta aggattatgt atatgacaaa 720
gacatagaac agattgccaa agaaacaaaa gaaaaaattt tgagattatt aacttccacc 780
tcaacagatg tttctcatca acagtttatg acatcagggt ttgaagacaa gcaatcaacc 840
tgtgagacaa aggaacagga gccaaaattg gtgaaaccca agaaaaagag aagaaaaaag 900
tcagtctata ctgtaggcct gagagggcta atcaatcttg ggaacacttg ttttatgaat 960
tgtattgtcc aggcacttac ccatattcct ctactgaaag atttcttcct ctctgacaag 1020
cacaaatgta taatgacaag ccccagcttg tgtctggtct gtgaaatgtc ttcgcttttt 1080
catgctatgt actctgggag ccgaactcct cacattccct ataagttact gcatctgata 1140
tggatccatg cagaacattt agcagggtac aggcagcagg atgcccatga gttccttatt 1200
gcaatattag acgtgctaca tagacacagc aaagatgata gtggtgggca ggaggccaat 1260
aaccccaact gctgtaactg catcatagac caaatcttta caggtggcct gcaatcagat 1320
gtcacatgtc aagcctgcca tagtgtttct accaccatag acccatgctg ggacatcagt 1380
ttggacttgc ctggctcttg tgccacattc gattcccaga acccagagag ggctgacagc 1440
acagtgagca gggatgacca cataccagga atcccctcac ttacagactg tctacagtgg 1500
tttacaaggc cagagcacct aggaagcagt gccaaaatca aatgcaatag ttgccaaagc 1560
taccaggagt ctactaaaca gctcacaatg aaaaaattac ccattgtggc ttgttttcat 1620
ctcaagcggt ttgagcatgt aggcaaacag aggcgaaaga ttaatacctt tatctccttt 1680
cccttggagc tggacatgac tccgtttttg gcctctacta aagagagcag aatgaaagaa 1740
ggccagccac caacagattg tgtgcccaat gagaataagt attccttgtt tgcagtgatt 1800
aatcaccatg gaactttgga aagtggccac tataccagct tcatccggca acaaaaggac 1860
cagtggttca gctgtgatga tgccatcatc accaaggcta ccattgagga cttactctac 1920
agtgaagggt atttactgtt ctatcacaaa cagggtctag agaaagacta gtcttaccag 1980
accacttact gaaaaaaaag taaatgatta ggcaaggatt ttgaagtgac acacagacct 2040
acttggaatg gacaatgaca gtaacaccta tgtgacagct agtatcttga tataaagaac 2100
ctattttagc atggcccatg ggtctgtcgg aagaaaaaaa tgaatactaa ccagtgacca 2160
ttcaacctta agaaatgggg agagggagaa gaggttgaaa atggtcacat aaagcataat 2220
gaaatgaaaa gaatgcttta ggtggggaca acgggagtag aagtgttctg atgctactct 2280
atgtcatttg tttttacaga aatatcttgt gaagtcaggg agtattcctt tatcagcaaa 2340
aacttcacaa ttggtgttcc agctgtggct gaccagctaa atagtttgaa agaaaaataa 2400
tattttaaaa taaagtttaa agagctttaa aagaaaaaca tttaaaaagg aaaaaatcat 2460
ttttaagatt ttaaaagaaa aaaactttta aatgttgaaa aaaatttaag ttgttatttt 2520
taaaagaaat attttaaaag ttaaaaataa ttttttaatt taaagaagtt tcagaatttt 2580
aaaaattaaa agcaaagaaa attaaattct taaagtttaa aaatgtaaaa taaattaagg 2640
aacaaggtta aaaatgaaag tttaccaaaa aaaggaagaa aatactgtta aaaattaaag 2700
ttagaaacaa aggaacatct taaaagtttc aaatgaagga ataatataaa tagatatttc 2760
aaaattaaag cataaaatat acgtatttaa aaagtgttaa caaaattact actataatga 2820
ttaagaaata aattttcaaa aatacagaat ggaatgcaat tcagatttta gagaaaagtt 2880
ttaaaagagg caagtttaga ataattcaag acaaaaagac aaaatgtgtt taaagacaaa 2940
aattgacaaa atacaggaag aaaatagaga cttgtaaaat aaaaagaacc ttagataagt 3000
tcaagagatt taaatgaaaa ctttaaatat ttaaataaag atttaaaaat ttaagctttt 3060
aaaaagaaaa acagttacat aaaaattgac cagtgaaaaa atgtgaaaga ttccagtaga 3120
aaacattatt aaaattaaca ggtttaagag gtctattntt ttatttaagc at 3172

<210> 42

<211> 1997
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473300CB1

<400> 42
ataatccacc catggaacca atctgaaaag aatgcagtca gaccctggac ccagtctgaa 60
ggtgatgttc tgcaaccttg gatctatgct gaaagcaata cagtcagact ctgggcccat 120
tctgaaactg ataaaataaa acaatatact gagcctgaat ctcaagcaat taggatgtgg 180
cctgaagagg atatgttcgc actttggtcc ccaacacaaa acgatgcagt ttggccatgg 240
acccaagtgg aatcacaaat gacccactcc tggacccaga atcaacttag tataaattac 300
ccttggactc agcatgtacc tgctgcaatc agaccatgga cttactctga aattcaaccc 360
tgcacccacc ctgaagccaa tacagtgata agatactggt tccagactca aatgagttca 420
ttaaatcctg ggaccaacct gaaactgaag tattccaaat ttggactatg ttgctcacac 480
aaagcctgtt tgggggtctc ttcacacgga cacgtgagac agtttgtata tttcagccct 540
ggactcagca aagagttact acaaatcgtt cgtggaccca ccctgaaacc caagcagaga 600
gactctggat caagcaggaa actgaagata gagacagatc ttcgttttac attcaaatga 660
ataaaggcag accatgggtt tatttgaaat atcaaatagt cggcgcctgg atccagcctg 720
aacttgatgt aattcactct tttatccagt ctgaaacctt cctattaaga ttctggccca 780
aggttctatc tccagtagtc aaaccatgga tcttgcttaa aggaagaaca ctcatatctt 840
ggatactgcc tgtaacccga gcagacactg gatccagtct gaagttcatc ttattgaatc 900
cttcggtgtt tttaaagccg gcaaaccatc tgagtacctg ggaccgcagg cacacgctac 960
tgcatctgga taatttttgtt gttgttgttc ttgctgttga aagtcctgga attgtgcaaa 1020
aacggcacct gagcatccta caagtcagca cttgtgccca attttggctc aagctgaatg 1080
aactcacttt ctgggtggag gccaagaaag ccatgtggat ggctgactat cagggagtga 1140
cacagtctag ctatgctccc tggtacaagc aagggcccat gactacctct gcttctatgt 1200
cccattcagt ctctacctct acaaatgctt cagcttttac ctccaccccct gcttctcttt 1260
ggccacactt ctctctgcca cagcctcaga gtaaggctca aaaacttggt agagatcaga 1320
tttatctgcg atatgccatg ccttggaagg ctgtcatcat catctgtggg agtcagatct 1380
gcagtggttc catagttggc agctcttgga ttctcacagc tgcccactgt gtcaggaaac 1440
tcagggatcc tgaagacact gctgtgatac tgggcctgag gcatcctggg gcaccactga 1500
gagttgtgaa ggtgtctacc attctgctgc atgagagatt ctggttggtg actgaggcag 1560
caagaaatat tctggaattg ctactcctcc acgatgtcca gactcccatt tggctcttat 1620
cactcttggg ctatctgagg aacctgaata gttcagaatg ctggctctct aggccacata 1680
ttgttacacc agctgtcctg cttagacacc cctgggcccc agggggaccg caacctcacc 1740
caggcactgg accactccca cagattcagg ctcagcagcc taacctgcaa atccatcatg 1800
tagctcagca ggacttcatc atttgtgacc ctggtccata tctgggccca agtcttgagc 1860
accatgtgtt tctgggctgg ctccccgcaa ccctgctcct gggacctagg cgcccacccc 1920
ctgctgccag ccatcccgaa ttagcagctg cgaagacatg gctctggccc ggaaaccggg 1980
gatgccctgt ggcttga 1997